

# **Moral Error Theory:**

## **A Cognitivist Realist Defence**

Wouter Floris Kalf

Submitted in accordance with the requirements for the degree of Doctor of  
Philosophy in Philosophy

The University of Leeds  
School of Philosophy, Religion and History of Science

August 2013

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Wouter Floris Kalf to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

© 2013 The University of Leeds and Wouter Floris Kalf

## Acknowledgements

First of all I would like to thank my supervisors at Leeds, Dr Pekka Väyrynen and Dr Andrew McGonigal for stimulating supervisions and helpful advice. Thanks as well to Dr Ulrike Heuer, Dr Gerald Lang, Dr Matthew Smith, Dr Tim Taylor, Professor John Divers and Professor Graeme Gooday.

Many thanks to audiences at the following events for instructive feedback on presentations of some of the arguments that feature in this thesis: *The Postgraduate Session*, 87<sup>th</sup> Joint Session of the Aristotelian Society and the Mind Association, University of Exeter (2013); *Understanding Value Graduate Conference*, University of Sheffield, UK (2013); *British Society for Ethical Theory (BSET) Annual Conference*, University of Stirling, UK (special thanks to my respondent Simon Kirchin) (2012); *5<sup>th</sup> Annual Arché-CSMN Philosophy Graduate Conference*, University of St. Andrews, UK (special thanks to my respondent Brian McElwee) (2011); *Third Annual Dutch Conference on Practical Philosophy*, University of Amsterdam, the Netherlands (2011); *University of Oslo & University of Leeds CMM-CSMN 'Knowledge and Agency Workshop'*, University of Leeds (2010); *Centre for Ethics and Metaethics Seminar Series*, University of Leeds (x4); *Postgraduate Seminar*, University of Leeds (x5). I thank the universities of Leeds and St. Andrews, the Analysis Trust, the Aristotelian Society, and the Mind Association for financial aid that made attending some of these events more feasible.

Part of chapter 4 has been published as Kalf, W.F., (2013), 'Moral Error Theory, Entailment and Presupposition', in: *Ethical Theory and Moral Practice*; **16(5)**: 923-37. I thank Pekka Väyrynen, Andrew McGonigal, Gerald Lang, Ulrike Heuer, John Divers, Ben Saunders, Michael Bench-Capon, Paul Ramshaw, Thomas Brouwer, Carl Baker, Tim Taylor, Bart Streumer, Kent Hurtig, and an anonymous referee for BSET for helpful comments on various drafts.

I would also like to mention the conferences and workshops at which I have presented my paper 'The *Tractatus Theologico-Politicus* and the Received View of Spinoza on Democracy' (now forthcoming in *Res Publica*): *ChiPhi Early Modern Philosophy Workshop*, University of Sheffield (2011); *1st CEGP Political Theory Graduate Conference*, LUISS Guido Carli University, Rome, Italy (2010); *3rd Biannual Philosophy Graduate Conference*, Central European University, Budapest, Hungary (2010); *Nederlands-Vlaamse Dag voor de Filosofie (Dutch-Flemish Philosophy Day)*, University of Tilburg, the Netherlands (2009); *Graduate Student Conference*, Claremont Graduate University, California, USA (2009). Thanks to *ChiPhi*, the Central European University, Claremont Graduate University and the University of Leeds for funding some of my expenses for some of these events.

Thanks to my fellow postgraduate students at the University of Leeds for making my time here productive and fun. Special thanks to Michael Bench-Capon, Sean McNally and Thomas Brouwer for being good housemates, to Becky Bowd and Paul Ramshaw for drinks, to Sean McNally and Sarah and Dani Adams for football, and to Richard Caves, Jon Banks, Jordan Bartol, Carl Baker, Carl Warom, David Cornell and Rob Pezet.

I am also grateful for financial support from a Marie Curie International Reintegration Grant funded by the European Community FP7/2007-2013 (fees and

maintenance in years 1-3 under grant agreement no. 231016) and the Jacobsen Studentship funded by the Royal Institute of Philosophy (year 4).

Finally I would like to thank my grandfather Carel Lodewijk Kalf for convincing me (but especially my parents) to go to '*Het Stedelijk Gymnasium*' in Schiedam to get a diploma with which I would be able to go to university. Without that advice I would not have been where I am now. Thanks to Laila Al-Jarml for proofreading. Most of all, however, I thank my father Carel Albert Kalf, my mother Veronica Elizabeth Maria Kalf-Müller and my brother Laurens Marnix Kalf for their unfailing support throughout my studies, first at Leiden University, then at University of California-Irvine and finally at the University of Leeds. Without them I would never have been able to achieve what I have.

## Abstract

This dissertation argues that moral error theory is the most plausible metaethical theory if we assume the truth of cognitivism about moral judgments and the moral statements that express them. According to moral error theory, various moral statements carry a non-negotiable commitment to a robust kind of categorical moral normativity, which means that this commitment cannot be denied on pains of changing the subject. Unfortunately, there is no such robust categorical moral normativity, at least not in the actual world. This entails that these moral statements are always untrue, or ‘in error’.

In arguing for moral error theory, the thesis first argues that the standard argumentative strategy for establishing moral discourse’s non-negotiable commitment—viz., forging a relation of conceptual entailment between moral statements and the statement that there exists robust categorical moral normativity—is highly problematic. It also argues that forging a presupposition relation can work, but that error theorists are best advised to pursue a completely new strategy, which uses a relation of *metaphysical* entailment. The dissertation then argues that moral discourse metaphysically entails robust moral categorical normativity and proceeds to present a new argument against its existence. According to this argument, various sorts of hypothetical *and* categorical normativity exist because these can be grounded in a naturalistically respectable metaphysic; unfortunately, categorical *moral* normativity cannot be so grounded.

Finally, the dissertation explores an often ignored answer to the following question: what (prudentially) should we with our error-riddled moral discourse? I argue for revolutionary cognitivism. This is the view that we should continue to use moral language and fully believe what we say but that *what* we say should be purged of its error. We should revolutionize our moral thought and start to conceive of morality’s normativity in a less robust way than we currently do.

## Table of Contents

<b>Acknowledgements.....</b>	<b>iii</b>
<b>Abstract.....</b>	<b>v</b>
<b>Table of Contents .....</b>	<b>vi</b>
<b>Chapter 1 Introduction: Moral Error Theory .....</b>	<b>1</b>
1.1 Introduction.....	1
1.2 The Metaethical Debate and Moral Error Theory.....	1
1.3 Rival Metaethical Theories .....	10
1.3.1 Cognitivist Theories.....	11
1.3.2 Non-cognitivist Theories .....	16
1.3.3 Further Theories .....	18
1.4 Problems for Moral Error Theory .....	20
1.5 Conclusion and Preview .....	22
<b>Chapter 2 Existing Error Theories .....</b>	<b>23</b>
2.1 Introduction.....	23
2.2 J.L. Mackie .....	23
2.2.1 Mackie's Early Moral Skepticism .....	23
2.2.2 Mackie's Non-Negotiable Commitment Claim .....	27
2.2.3 Mackie's Substantive Claim .....	38
2.2.4 Mackie's Solution .....	52
2.3 Richard Joyce .....	57
2.3.1 Joyce's Non-Negotiable Commitment Claim .....	57
2.3.2 Joyce's Substantive Claim .....	73
2.3.3 Joyce's Solution .....	81
2.4 Other Error Theories .....	84
2.5 Conclusion and Preview .....	90
<b>Chapter 3 Options for Error Theorists .....</b>	<b>91</b>
3.1 Introduction.....	91
3.2 The Non-Negotiable Commitment Claim.....	92
3.2.1 Entailment .....	93
3.2.1.1 Logical Entailment .....	94
3.2.1.2 Conceptual Entailment .....	95
3.2.1.3 Metaphysical Entailment .....	97

3.2.2 Implicature .....	99
3.2.2.1 Conversational Implicature .....	99
3.2.2.2 Conventional Implicature .....	100
3.2.3 Presupposition .....	103
3.3 The Substantive Thesis .....	107
3.4 Error-riddled Moral Discourse .....	109
3.5 After Moral Error Theory .....	110
3.6 Conclusion and Preview .....	112
<b>Chapter 4 Conceptual Entailment and Presupposition .....</b>	<b>113</b>
4.1 Introduction .....	113
4.2 A New Objection to the Entailment Claim .....	113
4.3 Entailment and Pragmatics .....	121
4.4 Against Conceptual Entailment .....	123
4.5 Against Presupposition .....	129
4.6 Conclusion and Preview .....	132
<b>Chapter 5 Metaphysical Entailment .....</b>	<b>133</b>
5.1 Introduction .....	133
5.2 Platitudes .....	134
5.3 The Metaphysical Entailment Claim .....	140
5.3.1 Categorical Moral Normativity .....	140
5.3.2 Categorical Moral Normativity in the Rationality-Involving Sense .....	146
5.4 Doing With Less .....	155
5.5 Conclusion and Preview .....	158
<b>Chapter 6 Categorical Moral Normativity in the Rationality-Involving Sense .....</b>	<b>159</b>
6.1 Introduction .....	159
6.2 Grounding Procedural+ Normativity .....	163
6.2.1 Functional Norms .....	163
6.2.2 Problems and Solutions .....	168
6.3 Norms of Prudence and Epistemology .....	174
6.3.1 Norms of Prudence .....	174
6.3.2 Norms of Epistemology .....	182
6.4 Procedural+ Categorical Moral Normativity? .....	187
6.4.1 Do Procedural+ Rational Requirements Entail Categorical Moral Requirements? .....	187

6.4.2 Can Categorical Moral Norms be Grounded in the Functional Profile of Human Psychologies? .....	195
6.5 Looking Elsewhere .....	198
6.6 Conclusion and Preview .....	209
<b>Chapter 7 Further Objections to Moral Error Theory .....</b>	<b>210</b>
7.1 Introduction .....	210
7.2 Quick Objections, Quick Replies .....	210
7.3 Honesty .....	212
7.4 Pervasiveness .....	214
7.5 The Non-Arbitrariness Constraint and Metaphysical Possibility .....	217
7.6 Conclusion and Preview .....	219
<b>Chapter 8 Revolutionary Cognitivism .....</b>	<b>220</b>
8.1 Introduction .....	220
8.2 Psychology, Semantics, Epistemology, Benefits .....	220
8.3 Objections and Replies .....	232
8.3.1 A Collapse Into Realism?.....	232
8.3.2 Feasibility .....	233
8.3.3 Doxastic Involuntarism .....	236
8.4 Comparing Revolutionary Cognitivism with Other Views .....	238
8.4.1 Abolitionism.....	238
8.4.2 Fictionalism.....	239
8.4.3 Preservatism .....	241
8.5 Conclusion .....	243
<b>Bibliography .....</b>	<b>244</b>



# Chapter 1

## Introduction: Moral Error Theory

### 1.1 Introduction

This chapter introduces moral error theory.<sup>1</sup> It explains the general metaethical debate and how error theory fits into that debate (§1.2). It also explores rival metaethical theories and explains how error theory responds to those theories (§1.3). Both of these sections introduce key terms and concepts that are used throughout the dissertation. The chapter continues to discuss problems for error theory (§1.4). Solving these problems is an important task for the remainder of the dissertation. A final section concludes and looks ahead (§1.5).

### 1.2 The Metaethical Debate and Moral Error Theory

The metaethical debate contrasts with but is related to the debate about normative ethics. Their relation lies in the fact that they are both elements of the *philosophical explanation of morality*, which is an explanation of which actions are right and wrong (normative component) as well as where morality comes from and why we should care about it (metaethical component).<sup>2</sup>

Which types of actions are included in the explanation? For my purposes in this and the next few chapters I will make use of the idea that we can ‘recognize the moral when we see it’.<sup>3</sup> For instance, we all recognize that causing an innocent child to undergo a prolonged and excruciatingly painful death is a morally relevant action

---

<sup>1</sup> I will often abbreviate moral error theory to ‘error theory’; unless I compare moral error theory with other kinds of error theory, such as error theory about mathematics (Field 1989) or colour (Ellis 2006), in which case I will write moral error theory.

<sup>2</sup> Egan (2007: 206). This distinguishes philosophical ethics from such related fields as sociological and theological ethics, which study, respectively, which moral rules people as a matter of fact accept or have accepted and which moral rules accord with a particular religion.

<sup>3</sup> Joyce (2001: 67, 75); Mackie (1977: 29); Brink (1986: 29). Similarly, Shafer-Landau (2003: 80) writes that we “know [the natural] when we see it” much like we can recognize “the pornographic” when we see it. Also see Papineau (2007).

(and indeed a moral kind of wrongness, if anything is).<sup>4</sup> By contrast, and other things being equal, we all recognize that believing that grass is red in the face of (known) conclusive evidence that grass is not red is not morally relevant (it is an epistemic kind of wrongness, if anything is). If other things aren't equal, such as when the devil has promised me that he will cause an innocent child to undergo a prolonged and excruciatingly painful death unless I believe that grass is red in the face of (known) conclusive evidence that grass is not red, then the question of whether to believe that grass is red becomes morally relevant (and the wrongness of not believing that grass is red will be moral and not epistemic).

I leave the content of morality vague at this part of the dissertation because a precise delineation of its boundaries requires careful discussion that is most naturally introduced in later chapters.<sup>5</sup> Moreover, that discussion would be distracting in the present chapter, which aims to explain and introduce error theory. As will become clear, this aim can be attained almost solely on the basis of reflections on the *status* of morality, thus without reflections on its *content* or *subject-matter*.

I continue with my explanation of the metaethical debate. Both metaethics and normative ethics partake in the philosophical explanation of morality. Their difference lies in the fact that, roughly speaking, normative ethical theories concern the first part of the philosophical explanation of morality (which actions are right and wrong?) whereas metaethical theories concern its second part (where does morality come from and why should we care about it?).<sup>6</sup>

---

<sup>4</sup> This example is due to Shafer-Landau (2003: 207). Copp (2007: 119) uses it as well. I will more often use the example of the wrongness of stealing for ease of exposition (and because I find Shafer-Landau's example distasteful). See Singer (1972: 231) for another example of an obvious moral truth; viz., that avoidable, undeserved human suffering is morally bad.

<sup>5</sup> One of many issues is that an 'other-regarding' conception of the content of moral norms is inconsistent with what some regard as our moral obligation not to kill *ourselves*. It is sometimes argued that the moral concerns other-regarding behaviour and that the ethical allows for more inclusive evaluations of agents, including their self-directed actions. I will however, *pace* Williams (1985a: 6) and Skorupski (1999: 1) but with Enoch (2011a: 2), use the terms 'moral' and 'ethical' interchangeably

<sup>6</sup> This is 'roughly speaking' because philosophers disagree about what, exactly, is included in ethical and metaethical theorizing (Enoch 2011a: 17, 41, 49). Moreover, some paradigmatically normative ethical theories like *rule consequentialism* can

Normative ethical theories enumerate which actions are right or wrong and explain why they are right or wrong.<sup>7</sup> One but by no means the only possible way in which philosophers proceed in this regard is by providing a Theory of the Good and a Theory of the Right. Hedonistic Utilitarians, for instance, might claim that what is good is pleasure and that what is right is to maximize pleasure. Armed with this theory, Hedonistic Utilitarians can enumerate which actions are right and wrong (because right actions maximize pleasure, wrong ones don't) and explain why that is so (because their theories of the Good and the Right are justified).

Metaethical theories, by contrast, are theories about the nature or status of moral reality itself, not about which action types or tokens are morally right or wrong and why. Metaethicists are concerned with the ontological status of moral truths, if indeed there are any (moral metaphysics).<sup>8</sup> They are also concerned with our epistemic access to these moral truths, again if there are any (moral epistemology) as well as with the nature of moral judgments (moral psychology) and moral statements (moral semantics), which undeniably exist.<sup>9</sup>

In an attempt to reach an overall theory of the status of morality, which includes an account of the nature of our judgments and statements about it,

---

also be understood as metaethical proposals about the nature of moral facts (Copp 2010: 142; cf. Enoch 2011a: 46-7, 47n60). I use this (simplified) understanding of normative ethics unapologetically as I use it solely for the purpose of introducing metaethics and through that error theory. What I think of as the correct relation between normative ethics and metaethics is one of my topics in the final chapter of this dissertation.

<sup>7</sup> Darwall et al. (1992: 181). Note that normative ethics deals with deontic concepts like RIGHT and WRONG, evaluative concepts such as GOOD and BAD, and thick concepts like BRAVE and COURAGEOUS (henceforth I use SMALL CAPS when I talk about concepts).

<sup>8</sup> Darwall et al. (1992: 126). Jackson (1998: 1-5) talks about the *location problem* for ethics.

<sup>9</sup> That is, it is undeniable that there are judgments about morally relevant matters, such as judgments about stealing and killing and it is also undeniable that some of these judgments become *statements* by being uttered or inscribed (what is deniable, for instance, is whether the moral terms in such statements make a separate contribution to their meaning—Ayer 1936). Note that an accidentally uttered string of words—e.g., by someone who has no knowledge of the English language—that sounds like 'stealing is morally wrong' is not a moral statement, nor is a similarly accidentally inscribed string of letters 'stealing is morally wrong' (Putnam 1992; Glassen 1959; Joyce 2001: 12-3). The list of metaethical topics given in the text is not meant to be exhaustive. For instance, Gill (2008: 99) adds moral phenomenology to this list. Also see Miller (2013: 2).

metaethicists take up two jobs. The first job is to present the most accurate account of moral discourse as possible.<sup>10</sup> Metaethicists must account for our actual moral discourse and practice in a way that does justice to its deeply entrenched features (perhaps shoring up any wrinkles or inconsistencies in the process). For example, if in making moral statements we are expressing pro- and con-attitudes towards actions which are not (correspondence) truth-apt, then this feature of moral discourse has to be part of our overall metaethical theory.<sup>11</sup> However, if instead what we are doing in making moral statements is trying to describe a mind-independent moral reality so that our statements are correspondence truth-apt then *that* has to be part of our metaethical theory.

The rationale behind doing this first job is that metaethicists want to make sure they are talking about, and ultimately vindicating or debunking, the real thing—*moral* discourse. They don't, and shouldn't, want to be talking about and ultimately vindicating or debunking something that looks a lot like moral discourse but on

---

<sup>10</sup> I define moral discourse as including not only ethical talk but also ethical beliefs, reactions, a sense of obligation and more besides (Hartcourt 2005: 251; Evans and Shah 2012: 85). Even so it is not entirely clear what 'ordinary moral discourse' or 'folk morality' *is*. Is it "unreflective bourgeois morality" (Nagel 1986: 166), an ideally cleaned-up version of that (Jackson 1998), different still? I present my own account in Chapter 5.

<sup>11</sup> I add 'correspondence' to truth aptitude because pro- and con-attitudes might be truth-apt in a more minimal way. This may happen for instance when *disciplined syntacticism* is true. According to disciplined syntacticism, so long as a sentence is suitably well-behaved—i.e., so long as it can be embedded in propositional attitude-contexts and plays an appropriate role in conditionals, logical inferences and the like—then it is truth-apt (Boghossian 1990: 161-7). Henceforth, for simplicity I'll sometimes talk about moral judgments and sentences being truth-apt although the notions of truth and falsity apply in the first instance to moral propositions. (Propositions are thought to be (any combination of): (i) the meanings of sentences, (ii) the referents of that-clauses, (iii) the (primary) bearers of truth-value, and (iv) the objects of beliefs and other 'propositional attitudes' such as hopes and fears; McGrath (2012: §1); Williamson (2007: 15). There are two main theories of propositions: propositions can be conceived of as complexes of senses or abstract entities (Fregean propositions) or as complexes of ordinary concrete objects that are the referents of our words (Russellian propositions). It will not matter for my purposes which view of propositions we shall keep in mind, as both are congenial to moral error theory. Russellian moral propositions are flawed if the referents of our words (moral facts and properties) don't exist; Fregean moral propositions are flawed when its senses have an empty extension.)

closer inspection in fact just isn't the real thing (call this *schmoral* discourse).<sup>12</sup>

Methodologically, metaethicists conduct quasi-empirical investigations into the features of moral discourse but also employ traditional analyses of moral concepts. Quasi-realist empirical investigations are investigations of the metaethicist's own experiences with and intuitions about moral discourse.<sup>13</sup> Such investigations by no means reach the currently accepted standards in the sciences for empirical research. But they are useful, and often accepted, so I will make use of them as well.

One way in which the metaethicist could go wrong in this regard is when she projects her pre-conceived or professionally invested theoretical commitments unto moral discourse. Another difficulty for the metaethicist is that moral discourse contains various aspects that are to various extents superficial, inessential, contradictory and misleading.<sup>14</sup> This is where the question of how inconsistencies in moral discourse are to be shored up becomes relevant. For what does the most accurate picture of moral discourse look like if moral discourse itself contains contradictory elements? I will have a lot to say about this in Chapter 5; here I continue with my rough explanation of metaethics and through that error theory.

The second task for metaethicists is to come up with a theory of morality's nature and status that accounts for these commitments of moral discourse and that is philosophically respectable. An important aspect of a theory's philosophical respectability is its ability to meet the constraints of *metaphysical naturalism*; the ontological doctrine according to which, roughly, the only facts and properties that exist are natural facts and properties.<sup>15</sup> Other constraints include the theory's

---

<sup>12</sup> Joyce (2007b, MS). The first reference to *schmoralizing* that I know of occurs in Blackburn (1984).

<sup>13</sup> See Johnston (1987); Robinson (2009: 321); Dunaway et al. (forthcoming); Marks (2013: 9); Finlay (2008b: 136) for a defence of this methodology.

<sup>14</sup> Anscombe (1958); Kirchin (2010).

<sup>15</sup> Or at least this is a common assumption; see, e.g., Timmons (1999); Scanlon (2009); Miller (2013). Influential accounts of what makes a property a natural property cite their causal efficacy (Parfit 2012 Vol2: 305-6) or the nature of our epistemic access to them (Copp 2012: 28-9). A different take on natural properties doesn't specify a feature that they all have in common (what, for example, is the causal efficacy of the property of belonging to a certain species, which probably counts as a natural property?). This is the disciplinary approach according to which the natural is defined in terms of the subject matter of various disciplines (Shafer-

parsimony and explanatory power.

Let's give these projects names. Call the first part of the metaethicist's job the project of *internal accommodation* (because it is concerned with questions internal to our moral practice). And call the second part the project of *external accommodation* (because it is concerned with questions external to our moral practice).<sup>16</sup>

The error theorist's distinctive metaethical claim can now be understood. It is that it is impossible to do full justice to both the internal and external accommodation projects within the following constraints. For moral discourse to be *moral* discourse, error theorists argue, we have to acknowledge the existence of a moral reality, sufficiently robustly understood, that is the truth-maker of our moral beliefs and moral statements.<sup>17</sup> To say that moral discourse could still be *moral* discourse if we denied its commitment to a robust moral reality that supplies the truth-makers for our moral beliefs would be to fail at the internal accommodation project. That would be to cease talking about moral discourse and to start talking about *schmoral* discourse. Unfortunately, error theorists continue, accepting this as an essential part of the internal accommodation project makes a successful external accommodation project *impossible*. For a robust moral reality would be inconsistent with metaphysical naturalism, or cannot be accounted for in a philosophically respectable way more generally.

I have so far described moral error theory as taking ordinary moral discourse as its object. There is a purely explanatory reason for doing that in this introductory chapter: moral error theories have traditionally been formulated with ordinary moral

---

Landau 2003: 58). I present my own take on metaphysical naturalism in Chapter 6. Also, I understand facts to register the instantiation of properties (Shafer-Landau 2003: 65; Horgan and Timmons 2008b: 270-3); natural facts register the instantiation of natural properties and moral facts, if they exist, register the instantiation of moral properties.

<sup>16</sup> Taken, with adaptations, from Timmons (1999: 12); also see Southwood (2010: 8-9); Nichols (2004: 168). Not all metaethical theories have been constructed with these desiderata in mind; Timmons mentions McDowell (1985) and Platts (1991).

<sup>17</sup> Exactly how we should understand this notion of robustness is a difficult matter, and will be topic of much that is to follow, starting with a discussion of Mackie's and Joyce's versions of moral error theory in Chapter 2.

discourse as the object of their scepticism.<sup>18</sup> But there is a justifying reason as well. This is that we have to get our account of what moral properties are like *from somewhere*. After all, a metaethical account of morality according to which moral goodness is analysed as something that is “pink with yellow trimmings” so that moral goodness can be argued to exist because there *are* pink things with yellow trimmings, it will be agreed on all sides, cannot be accepted as a vindication of moral goodness.<sup>19</sup> The question is though: wherefrom do we get our account of what morality would have to look like to *count* as morality if not from the way we think and talk about morality; viz., from moral psychology and moral language?<sup>20</sup> The difficulty of answering that question explains why error theories take moral discourse as the object of their scepticism.

This reasoning notwithstanding, there are error theories that focus on something other than ordinary moral discourse. First, we can formulate an error theory about morality itself, either conceived of as a collection of entities including ought facts and moral properties, or as a system of rules, or again as a doctrine or ideology.<sup>21</sup> When error theorists focus on morality itself as opposed to moral discourse they have to be careful how they formulate their theory. For there is a good sense in which it is undeniable that various such moralities exist, at least when

---

<sup>18</sup> See Mackie (1977). One explanatory hypothesis for this fact is that Mackie worked in an era in which philosophy of language was considered to be the *first philosophy* (Burge 1992). So although his moral error theory reacts to that idea in a way by giving metaphysics more weight than the ordinary language philosophers concerned with metaethics before him, he still found it natural to *start* with a careful analysis of moral language couched in moral statements. Also, note that the term scepticism has at least two senses in contemporary philosophy. The first is one that denotes a mental stance of *withholding judgment* on an issue because one finds the considerations for two mutually inconsistent claims about an issue to be equally compelling (Pyrrhonian scepticism). The other sense of ‘scepticism’ is one that denotes a mental stance of *positively rejecting* a certain claim because it lacks sufficient justification (Cartesian scepticism). In this thesis ‘scepticism’ will be understood in the latter sense of the term.

<sup>19</sup> Stevenson (1937: 14). The implausible idea that words can mean whatever we wish for them to mean is known as *Humpty Dumpty’s Dictum* (Jackson 1998: 118; Schroeder 2007: 73-5). Also see Joyce on defining reasons as creatures that live in the forest (2001: 87) and Copp on defining heat as a cabbage or a king (2012: 47).

<sup>20</sup> Wedgwood (2001: 1-4).

<sup>21</sup> Clark (2009: 204); Cuneo (2012: 126-30).

these are understood as *not epistemically endorsed*.<sup>22</sup> That is, the error theorist should not deny that there exist moralities such as Christian Ethics and Moorean Intuitionism, as long as is what is meant by that claim is that what is being quantified over are people's false beliefs and patterns of behaviour. What such error theorists should (and do) deny is that such beliefs *should be endorsed* (where the 'should' is a non-moral, epistemic should to avoid self-refutation). This complication comes in addition to that which such error theorists face regarding the issue of how they are going to delineate their subject matter. They do not get their account of what morality looks like from considerations of the workings of moral language and moral psychology; but then where do they get this account from?

A further possibility for an error theory that doesn't focus on ordinary moral discourse is to formulate an error theory about *philosophers'* conceptions of morality.<sup>23</sup> A final possibility is an "amalgam or fusion of [philosophical and folk conceptions of morality]."<sup>24</sup> In what follows, for ease of exposition and following the current trend in the literature, I will take an error theory of folk moral discourse as my main example of moral error theory.

To sum up the discussion so far, moral error theory maintains that moral discourse is in error because it carries an essential or *non-negotiable* commitment to a robust kind of moral reality, the denial of which would change the subject from moral to *schmoral* discourse.<sup>25</sup> It also maintains that this commitment is flawed, for example because it violates metaphysical naturalism.<sup>26</sup> Since moral discourse consists of various sorts of moral judgments and statements, including, but certainly not limited to, 'stealing is morally wrong' and 'keeping promises is morally obligatory', moral judgments and statements too are flawed. This gets us the

---

<sup>22</sup> Joyce (2007b: 64).

<sup>23</sup> Mackie (1977: 35). This option was also on Hare's mind when he wrote that although many *philosophers* "have thought that moral words connoted [a robustly understood moral reality]" he didn't believe that "ordinary people, innocent of any philosophy, are the whole time committing the same error" (1981: 78-9).

<sup>24</sup> Blackburn (2005: 327).

<sup>25</sup> The term 'non-negotiable commitment' is Joyce's (2001: 4, 97).

<sup>26</sup> I use the term 'flawed' as a catch-all phrase for all the different ways in which moral discourse, or the statements that comprise it, can be 'in error': so moral statements can be false, untrue, neither true nor false, etc. I canvass all the options in Chapter 3.



following preliminary formulation of moral error theory:

**Moral Error Theory**

**P1 Non-Negotiable Commitment Claim:**

Moral statements and judgments carry a *non-negotiable commitment* to a robustly understood moral reality

**P2 Substantive Claim:**

This commitment is flawed

**P3 Auxiliary Claim:**

If moral discourse carries a false non-negotiable commitment then it is itself flawed<sup>27</sup>

**C Conclusion**

Moral discourse is flawed: the statements and judgments comprising it are 'in error' [From P1, P2, P3]<sup>28</sup>

Almost all contemporary moral error theories operate in much like this way.<sup>29</sup> In the remainder of this chapter I discuss the main rivals of error theory

---

<sup>27</sup> In contrast with almost everyone else in the field, Evans and Shah (2012: 86n13) argue that P3 is required for the argument to go through. I agree. If we think that the non-negotiable commitment of moral discourse is to a robustly understood notion of a moral reality then perhaps moral statements cannot be correctly asserted *relative to the 'correspondence-truth norm'*, according to which moral judgments are correctly assertible if and only if there exist moral properties that are the truth-makers of these judgments. But that doesn't rule out the possibility that they can be correctly asserted relative to the *'norm of usefulness'*, according to which moral judgments are correctly assertible if and only if they are useful (for instance, because they facilitate human cooperation—see Mackie 1977: Ch.5). For moral judgments can still be useful even when they are false (Timmons 1999: 79; Wright 1996: 3; 1922: 10; also see Miller 2002). Auxiliary Claim saves Conclusion because it rules out this possibility.

<sup>28</sup> Error theories are sometimes presented with a fourth step; the so-called explanatory step, which aims to explain why we hold the moral beliefs that we do and why we have had them for so long even though they are also false. Mackie employs his projectivist theory of morality, whereas Joyce appeals to our evolutionary history; Mackie (1977: 42); Joyce (2001: Ch.6, 2006). I argue that this step is logically superfluous for Mackie's error theory in §2.2.3. The same holds, by implication, for Joyce's theory and my own.

<sup>29</sup> Different ways of arguing for error theory will be discussed in the next chapter. Some moral error theorists I won't discuss in much detail (but which are worth mentioning to get a sense of the presence of the theory in the literature) include Hinckfuss (1987), Biehl (2008); Garner (1994, 2010); Burgess (2010); Pigden (2010); Williams (1985a); Marks (2013). Historical figures that can be interpreted as error theorists *avant la lettre* include: Ramsey (1990); Hume (1739-40, 1751)

(§1.3). I also explain the main objections to error theory (§1.4). The remainder of the dissertation aims to show that an error theory that undertakes relevantly different commitments from the existing error theories I discuss in the next chapter can avoid many of these objections. It will also argue that my formulation of error theory can handle objections that apply specifically to it. This amounts to a sustained defence of error theory.

Before doing this, however, there is one terminological issue to resolve. Some philosophers regard theories according to which there *are* moral facts and properties and according to which it is just that the folk have *false beliefs* about these properties as “moderate” error theories.<sup>30</sup> I regard these as *success* theories: there *are* moral facts and properties after all, it is just that the folk have false beliefs about what those properties are.<sup>31</sup> I don’t want to imply that such theories aren’t an interesting addition to the metaethical landscape. I just won’t be concerned with them in this thesis because I think that the folk’s mistake about morality is stronger than this.

### 1.3 Rival Metaethical Theories

Typically, error theorists are *cognitivists* about moral judgments, believing that moral judgments are or express beliefs and that moral statements are truth-apt.<sup>32</sup> Error theories are also anti-realists about (the subject-matter of) those moral judgments. Hence the main rivals of error theory are cognitivist *realists* (§1.3.1) and *non-cognitivist* anti-realists (§1.3.2). Recent work in metaethics has also generated contenders that don’t fit nicely into either of these camps. An example of such a theory will be discussed as well (§1.3.3).

---

according to Olson (2011c); Nietzsche (1883-88) according to Pigden (2010: 20); Russell (1922); Protagoras according to Hare (1999). Smith (2006) has offered arguments for moral error theory without embracing it. Hussain (2010: 342-4) formulates a non-cognitivist version of moral error theory.

<sup>30</sup> Olson (MS: 14); also see Kahane (2013: 150); Miller (2009); cf. Shafer-Landau (2003: 17); Hussain (2010).

<sup>31</sup> The term ‘success theory’ was first introduced into the debate by Sayre-McCord (1986).

<sup>32</sup> See, e.g., Schroeder (2010: 12-5, 33-4); Miller (2013: 104).

### 1.3.1 Cognitivist Theories

Cognitivist theories hold that a moral judgment like ‘stealing is wrong’ expresses a (usually correspondence) truth-apt proposition about the nature of moral reality. Moral judgments either get the value ‘true’ or ‘false’ depending on whether or not they aptly describe this moral reality. As cognitivists that are not error theorists hold that at least some moral judgments are true, the only cognitivist theories that I will discuss in this sub-section are cognitivist realist theories.

Some cognitivist realists believe that moral judgments are correspondence truth-apt and that their truth-makers are supplied by a non-naturalist moral metaphysic.<sup>33</sup> Typically these theories are completely *non-revisionary* with respect to the internal accommodation project, at least by the error theorist’s lights. For non-naturalist realist theories can claim that the non-negotiable commitment of moral discourse is to a robustly understood mind-independent moral reality. After all, their outright acceptance of a non-naturalist moral metaphysic allows them to fully vindicate this picture of the status of morality in the external accommodation project. One reason for thinking that a non-naturalist metaphysic can be accepted is that it makes the best *overall sense* of our moral practice. A theory like error theory that insists on metaphysical naturalism and denies that morality exists loses more ‘plausibility points’ than a theory that requires us to enlarge our ontology with non-natural moral properties but is able to retain moral reality.<sup>34</sup> The error theorist’s response to non-naturalist moral realism, though, is to argue that non-naturalists are wrong to think that such a project can succeed.

There are also naturalist cognitivist realists. They too think that moral truth requires truth-makers for correspondence truth-apt moral judgments but they differ with non-naturalists about the nature of those truth-makers. For they think that moral reality doesn’t require a non-naturalist metaphysic. The only things we need are natural facts and properties.

Naturalist cognitivist realists divide themselves on at least three issues. The first is the Analytic/Synthetic issue:

---

<sup>33</sup> Enoch (2011a); Wedgwood (2007); Moore (1903); Shafer-Landau (2003) on some interpretations; Heumer (2005); also see Zangwill (2012: 347).

<sup>34</sup> Enoch (2011a: Ch.10); also see Shafer-Landau (2003: 43).

### **Analytic/Synthetic**

Are moral properties identical or else somehow fully constituted by natural properties in virtue of a metaphysical or a semantic relation?<sup>35</sup>

Analytic naturalists think that the second scenario obtains.<sup>36</sup> They often argue that the fact that moral predicates are semantically equivalent to predicates and sentences framed in non-moral (natural) terms do carry implications for the metaphysics of moral properties and facts. Standard defences of analytic naturalism assume the *synonymy test for property identity*: if two predicates are synonymous then the properties they stand for are identical (or, on a weaker formulation, that the two properties stand in a relation of constitution or something relevantly like a constitution relation).<sup>37</sup> Synthetic naturalists believe that moral properties are identical to or wholly constituted by natural properties because moral properties themselves are found, by *a posteriori* rather than *a priori* means, to be identical to or wholly constituted by non-moral (and in this case natural) properties.<sup>38</sup>

The second issue is about the reducibility of moral facts and properties:

### **Reductive/Non-Reductive**

Are moral facts reducible to non-moral (natural) facts, or are they not reducible to non-moral (natural) facts and do they exist in their own right?<sup>39</sup>

---

<sup>35</sup> Brink (1989: 162); Jackson and Pettit (1995: 28); Nuccetelli and Seay (2012: 131).

<sup>36</sup> Analytic naturalists include Spencer and Westermarck according to Moore (1903: section 29, 97-8; 1922: 332); Bentham (1789: Ch.1); Perry (1926); Sharp (1928); Jackson (1998). Note that there is substantial disagreement about “what portions of sentences [we should] account predicates” (Armstrong 1978: 2): predicate words such as ‘wrong’ and ‘circular’, expansions of these to include connective verbs to get ‘is wrong’ and ‘is circular’, or sentence-frames like ‘\_\_\_\_ is circular’. In this thesis I will write in accordance with the first option, but none of my argumentative purposes will be affected if one prefers to keep another option in mind.

<sup>37</sup> Brink (1989).

<sup>38</sup> Note that although “moral properties [may be] natural properties” synthetic naturalists often claim that we may not have “non-moral vocabulary [for them]” (Olson forthcoming-a: 29). This claim is often denied by analytic moral naturalism (Jackson 1998), and constitutes another difference between analytic and synthetic naturalism.

<sup>39</sup> Another way to put the point is to say that reductive theories are looking for type-type property identities whereas non-reductive theories are after something less than

Generally speaking, reductionist claims can be more or less reductionist in substance. For reductionist claims concerning entities of a kind K imply that Ks exist, albeit it that, contrary to the appearances, Ks are really Ys, where Ys are less ontologically problematic than and different from what we commonsensically assume Ks to be. The degree of departure from the commonsensical understanding of what Ks are, by the proposed reduction-base that are the Ys, is what determines how reductionist ‘in substance’ the reduction is. For an example consider psychological behaviourism.<sup>40</sup> Behaviourists are wary of beliefs *understood as mental states* (Ks). But they don’t deny that beliefs exist. Instead they reduce beliefs to behavioural dispositions (Ys). The reduction succeeds, claims the behaviourist, because enough of our commonsensical convictions about beliefs understood as mental states are retained by understanding them as behavioural dispositions. This includes, but is not limited to, the conviction that beliefs reliably cause us to act in certain ways, and behavioural dispositions do exactly that.

An example of a reductive view in metaethics is Jackson’s moral functionalism, according to which moral properties fully reduce to non-moral (or, as he prefers, descriptive) properties because of the specific relation of meaning-equivalence between moral and descriptive predicates.<sup>41</sup> An example of a non-reductive view is Sturgeon’s moral naturalism according to which moral facts and properties feature in our best explanations of the world and are therefore naturalistically respectable even though they are not reducible to non-moral (natural) properties.<sup>42</sup>

The reductionism/non-reductionism issue is orthogonal to the analytic/synthetic issue. A stance on the latter issue implies nothing about whether or not one is or should be a reductionist about the nature of moral properties. Synthetic naturalists are free to claim that their *a posteriori* investigations into the nature of moral properties reveals either that they do or don’t reduce to non-moral (natural) properties. Likewise, analytic naturalists can claim that correctly carried

---

that, such as actual world-instantiation, constitution, realization or tokening of moral property types by non-moral, natural features (Bedke 2012: 111).

<sup>40</sup> Cuneo (2007a: 29-30).

<sup>41</sup> Jackson (1998).

<sup>42</sup> Sturgeon (1984).

out analyses of moral concepts entail either that moral properties reduce to non-moral (natural) properties or that they don't.

The final issue I want to discuss here is:

### **Revisionist/Non-Revisionist**

Is the naturalist account of moral discourse and the nature of moral reality fully consistent with our honest results in the internal accommodation project or do they imply *benign* revisions of our conception of morality?<sup>43</sup>

Non-revisionist naturalist realists claim that *all* of the deeply embedded features of ordinary moral discourse can be accounted for with naturalistically (or more generally philosophically) respectable means.<sup>44</sup> Revisionists claim that although a naturalistically respectable moral metaphysics cannot account for *all* of the important deeply embedded features of ordinary moral discourse, it can account for *enough* of these so that *a kind of morality that doesn't collapse into a schmorality* is vindicated after all. In that case the revision is said to be 'benign'. For instance, revisionist naturalist realists might claim that almost all of the deeply embedded features of moral discourse can be accounted for purely by reference to natural properties, such as morality's content and motivational force, but that one such element, such as morality's normativity, cannot be so accounted for. Often the argument for a benign revision is that the overall philosophical picture of morality that entails that morality is real albeit slightly different from how we ordinarily think about it is more plausible than an overall philosophical picture of morality according to which the failure of the instantiation of the folk conception of morality commits us to metaethical nihilism. Lewis claims that a choice between these two accounts may be merely a "matter of temperament."<sup>45</sup> Railton's revisionist naturalism is more firmly grounded in a comparison of 'plausibility points'.<sup>46</sup>

How do error theorists argue against naturalist realists? Against analytic naturalists, regardless of whether or not they also accept the semantic test for

---

<sup>43</sup> Of course there are non-benign revisions, but these would give rise to error theories, so I won't discuss them here.

<sup>44</sup> Copp (2007, 2010).

<sup>45</sup> Lewis (1989: 137). Also see Lewis (1996).

<sup>46</sup> Railton (1986).

property identity, they argue that a correctly carried out conceptual analysis reveals that conceptual *non-naturalism* rather than conceptual naturalism is true. Moral predicates are not identical to natural predicates—it is at least also part of moral concepts that they involve non-naturalist, robustly normative aspects (this reply assumes that the error theorist has in hand an account what it takes for a predicate to be non-natural and robustly normative). Against the synthetic naturalist error theorists argue that their account of moral properties that denies that they would have to be part of a robustly understood moral reality ends up being an account of *schmorality* and that any attempt to account for that kind of normativity whilst remaining faithful to metaphysical naturalism will be impossible. Against *revisionary* naturalist realism the error theory argues that her proposed benign revision of morality is in fact an account of *schmorality* only.<sup>47</sup> The reducibility vs. non-reducibility issue is not of major importance to the error theorist.<sup>48</sup>

In addition to non-naturalist and naturalist realist theories, according to which moral judgements are true, when they are true, in virtue of correspondence with a non-naturalist or naturalist moral reality, there is a third type of cognitivist theories; *irrealist* cognitivism.<sup>49</sup> According to it, moral judgments express beliefs, are truth-apt, and sometimes true, but not in virtue of correspondence with *worldly* facts, whether natural or non-natural. Instead, moral judgments are true in virtue of *nominal* facts. (This also means that the nomenclature of ‘irrealism’ is a bit confusing, as moral facts are ‘real’ in *a* sense according to the irrealist; they are only not real in the same sense in which classical naturalists and non-naturalists think moral facts are real).

This view is a dangerous competitor to error theory as it insists, with error theory, on the non-arbitrariness or objectivity of moral facts without taking that to require any sort of metaphysical import whatsoever (this is the work that nominal facts are doing in this theory). However, error theory insists in turn that this project

---

<sup>47</sup> Revised moral concepts are like “*Hamlet* without the prince” (Joyce 2007a: §5) because the “[original moral] concept *is the whole point of the discourse ... it furnishes us with something authoritative that the wimpified [revised] surrogate [concept] cannot*” (Joyce MS: 14). Also see Jackson (1998: 45).

<sup>48</sup> There are many other objections to naturalist cognitivist realism, such as Fine’s (2002: 253) argument that naturalist theories entail the collapse of various kinds of necessity that should be kept distinct.

<sup>49</sup> Skorupski (1999, 2010).

must fail; that only a moral metaphysic existing at least partially outside of us and that we aim to be responsive to when we make moral judgments can account for the deeply embedded features of folk moral discourse. My own argument against irrealist cognitivism can be found in §6.5.

A final category of cognitivist theories that error theorists have to respond to is that of constructivism. According to constructivism there exist moral facts in virtue of ‘constructive’ efforts of moral agents. Moral facts are worldly facts alright, but unlike classical naturalist and non-naturalist theories they are also mind-dependent.<sup>50</sup> Examples are contractualist and contractarian theories.<sup>51</sup> Contractarianism is associated with Gauthier’s self-interested theory of agreements and contracts.<sup>52</sup> Contractualism is associated with Scanlon’s theory of agreements and contracts that is rooted in a standing desire to be moral that can come apart from whether moral behaviour is (always or usually) in the agent’s self-interest.<sup>53</sup> Since constructivists believe that moral facts are in an important respect mind-dependent, error theorists can respond by insisting that the mind-dependence of moral facts strays too far from the picture of morality as we get it from a fairly executed internal accommodation project.

### 1.3.2 Non-cognitivist Theories

In addition to cognitivist rivals error theory also has non-cognitivist rivals. These will now be discussed.

A radical non-cognitivist theory is Ayer’s, according to which moral judgments take a non-cognitive psychology and according to which a moral statement like ‘stealing is morally wrong’ *means* something like ‘boo! stealing’.<sup>54</sup> In

---

<sup>50</sup> O’Neill (1988/9: 3); Timmons (1999: 72); James (2007: 302); Heathwood (2012: 2); Shafer-Landau (2003: 14).

<sup>51</sup> Scanlon (2009, Lecture 4: 16) argues that contractualist theories are examples of constructivist theories only if the term ‘constructivism’ is taken in a sufficiently broad sense. Shafer-Landau and Cuneo (2007: 81) also puts Harman’s moral relativism down as a constructivist metaethical theory, as it holds that moral properties are conferred on things by actual moral agents. Harman views morality “as a compromise based on implicit bargaining” between agents (1975: 13).

<sup>52</sup> Gauthier (1986).

<sup>53</sup> Scanlon (1998). This distinction between contractualism and contractarianism is Darwall’s (2002).

<sup>54</sup> Ayer (1936); also see Stevenson (1937). See Joyce (2001: 201) for discussion.



making this claim Ayer makes the following transition from psychology to semantics. *Because* moral judgments aren't beliefs (psychological claim) and *because* it is this non-cognitive attitude that gets expressed by our public moral utterances (positive semantic claim), a further, negative semantic claim follows. This is the claim that moral predicates are only *syntactically* predicates; they don't function to pick out a property and their only semantic function is to indicate that a certain desire-state or emotion is present.<sup>55</sup> As we will see, not making this transition from psychological to particular semantic claims opens up logical space for metaethical theories that do not fit in the 'cognitivist realist' vs. 'non-cognitivist anti-realist' dichotomy (§1.3.3). Here however I focus on traditional non-cognitivist anti-realist theories.

The non-cognitivist camp also includes less radical, quasi-realist non-cognitivist theories such as Blackburn's.<sup>56</sup> Their view is 'less radical' because they try to 'earn the right' to talk about moral truth, moral reality, *et cetera*, even though notions of moral truth, moral reality, *et cetera* play no *fundamental* role in the explanation of moral judgments. One way in which they try to accomplish this task whilst steering clear of attaching metaphysical commitments to ordinary moral talk is by taking "what seems like a thought that embodies a particular second-order metaphysic of morals ... as a kind of thought that expresses a first-order attitude."<sup>57</sup> For by doing that the claim 'even if we had approved of it or desired to do it, bear-baiting would still have been wrong', although it sounds like a second-order realist commitment, can come out as a "perfectly sensible first-order commitment to the effect that it is not our enjoyments or approval to which you should look in discovering whether bear-baiting is wrong (it is at least mainly the effect on the bear)."<sup>58</sup> This then gets analysed in the familiar quasi-realist way. In trying to account for moral *truth* such theories can use a minimalist instead of correspondence theory of truth which similarly avoids ontological commitments.<sup>59</sup>

---

<sup>55</sup> Kalderon (2008: 1). Joyce (2006: 53) talks about non-cognitivism "presented as" psychological and semantic theses. Also see Joyce (2001: 8); Miller (2013: 112).

<sup>56</sup> Blackburn (1992: 948, 1993); also see Gibbard (1990, 2003).

<sup>57</sup> Blackburn (1993: 153).

<sup>58</sup> Blackburn (1993: 153).

<sup>59</sup> Minimalists hold that the question 'is *p* true?' can be answered simply by replying that *p* and that this reply incurs no substantial commitments about the

In this thesis I assume that non-cognitivism is false.<sup>60</sup> My project is to carve out the most plausible metaethical theory within a cognitivist framework.

### 1.3.3 Further Theories

There are additional metaethical theories that do not fit into my ‘cognitivist realism vs. non-cognitivist anti-realism’ dichotomy. Their grid in logical space can be secured by making the following distinctions:

<b>Cognitivism/Non-cognitivism</b>	Moral judgments take cognitive/non-cognitive psychology
<b>Expressivism/Anti-Expressivism</b>	The content of moral statements is/is not determined by the non-cognitive attitude it expresses
<b>Factualism/Non-Factualism</b>	The content of moral statements represents/fails to represent a distinctive domain of moral fact <sup>61</sup>

Classical non-cognitivists like Ayer have reasoned as follows. They accept the Non-Cognitivist thesis that moral judgments are non-cognitive attitudes and conclude that these attitudes get *expressed* in publicly available moral statements (that is, they move from Non-cognitivism to Expressivism). Moreover, because they accept both Non-Cognitivism and Expressivism they conclude from the claim that the content of moral statements is determined by the non-cognitive attitude it expresses that moral statements *fail to represent* moral reality (that is, they move from Non-cognitivism and Expressivism to Non-Factualism).

---

nature of truth. Correspondence theorists hold ‘is *p* true?’ cannot be ‘simply’ answered by replying that *p* and that this reply, when given, incurs the substantial commitment that there exists a state of affairs that answers to ‘*p*’ (Timmons 1999: 148-9; Wedgwood 2007: 6-7; Schroeder 2010: 26-30).

<sup>60</sup> Olson (MS) makes the same assumption in his forthcoming book-length defence of error theory.

<sup>61</sup> Kalderon (2005: vii-viii, 2008: 1).

However Kalderon has argued that these three theses can come apart. He defends *hermeneutic moral fictionalism*, which combines Non-Cognitivism with the semantic theses Expressivism and *Factualism*. Thus his theory maintains that moral statements are non-cognitive attitudes (they are attitudes of ‘make-believe’) and that although the content of moral statements is determined by the non-cognitive attitude it expresses, moral statements nonetheless express propositions which attribute moral properties to a state of affairs.<sup>62</sup>

This basic setup allows Kalderon to avoid error theory. He argues that moral propositions have both real and fictional content. The former (e.g., ‘this is an act of deliberate cruelty against children’) is completely naturalistically respectable and for people with the right moral sensibility sets the parameters for the fictional content of the moral proposition (e.g., ‘this act of deliberate cruelty against children is morally wrong’). Furthermore, the fictional content of a moral proposition solely has norms of correctness that are internal to the moral practice. Hence the real content of moral propositions requires a traditional inquiry into matters of correspondence with states of affairs in an empirical world and is completely naturalistically respectable, and the fictional content of the moral propositions does not require a traditional inquiry into matters of correspondence with states of affairs in an empirical world at all. Hence there isn’t anything objectionable about Kalderon’s theory from the point of view of a naturalistic metaphysics. Moreover, since at least some of our moral judgments come out as true, error theory is false. To this error theory responds by insisting that a non-cognitivist construal of the notion of moral acceptance entails a failure on the score of the internal accommodation project.

This concludes my discussion of the opponents of moral error theory. We have seen that there exists a rich variety of metaethical theories and I have had to ignore many of them to keep the discussion manageable. So advocates of a moral error theory have to be careful with how they formulate their theory. Ignoring distinctions and theoretical possibilities might play into the hands of the *success theorist*.

In fact, this is not the only problem for error theory. The next section surveys the main objections to error theory, and the chapter closes with a summary and

---

<sup>62</sup> Kalderon (2005: 149).

preview.

## 1.4 Problems for Moral Error Theory

Error theory has attractive and unattractive features. Its attractiveness lies in its honesty regarding the internal and external accommodation projects—no half-way house compromises are made. What is unattractive about error theory is its skepticism about morality. But what are the main philosophical problems with error theory that go beyond this initial observation?

One problem is the Formulation Dilemma:

### Formulation Dilemma

Error theorists can either formulate a *global* error theory via a broad meta-normative scepticism, or they can formulate a *local moral* error theory which leaves intact at least some normativity (and only finds, e.g., categorical moral reasons of rationality objectionable)

On the first horn, the problem is that the global meta-normative error theory is even more radical than the local moral error theory, and to that extent harder to defend. For according to the global meta-normative error theory, there are no normative epistemic facts, no facts about the correctness of logical inferences, no fact of the matter that if one wants to be in London by 5pm one has to take the 2.53pm train from Leeds at the latest, *et cetera*. On the second horn, the main difficulties for the local error theory are provided by slippery slope and entailment objections. Slippery slope objections are to the effect that if the local error theorist *accepts* *x* (where *x* could be hypothetical normativity) there is a good chance that she would have to be committed to *y* as well (where *y* could be categorical normativity) and where *y* is incompatible with her theory.<sup>63</sup> Entailment objections claim more strongly that denying *p* (where *p* is categorical moral normativity) *entails* denying *q* (where *q* could be categorical epistemic normativity), but that *q* is undeniable, so that a *reductio* ensues.<sup>64</sup>

A second problem is that moral discourse can't be said to carry any clear commitment whatsoever. Opponents of error theory have argued this point for a

---

<sup>63</sup> Korsgaard (1996); Hampton (1998). Also see Lillehammer (2007).

<sup>64</sup> Rowland (2013); Streumer (forthcoming). Also see Lillehammer (2007).

variety of reasons. One reason is that they think that moral discourse is too *messy* to carry a clearly formulated commitment. For every contender commitment we can find evidence in moral discourse that neutralizes it.<sup>65</sup> A second reason for thinking that moral discourse doesn't carry any essential commitments is that the commitments that error theorists single out are exaggerated:

ordinary people when they use [moral] words are not intending to ascribe objective prescriptive properties to actions<sup>66</sup>

I just do not think that ordinary moral discourse presupposes that ethics is objective *in the sense that realism and some versions of constructivism attempt to capture*<sup>67</sup>

it is extremely unlikely that any belief so *recherché* [as the error theorist's Non-Negotiable Commitment Claim] could be part of common sense moral thinking<sup>68</sup>

Both the first and the second problem concern the error theorist's Non-Negotiable Commitment Claim. A third problem for error theory concerns its Substantive Claim. It is that it is not at all clear that the mode of commitment of moral discourse is 'flawed'. This problem takes different forms depending on how error theorists argue the commitment is flawed. For instance, error theorists might argue that the commitment is flawed because it is inconsistent with our best theories about physics or ontology more broadly. In that case, the problem is at least threefold. First, 'scientism'—the view that the investigative methods of the physical sciences are applicable or justifiable in all fields of inquiry—it is agreed on almost all sides (and correctly so, I will assume), should be avoided, at least when our investigations are concerned with ethics.<sup>69</sup> But then we open up the possibility that there are methods of reasoning applicable or justifiable in the moral domain that vindicate moral facts even if we don't pay much attention to ontology. Second, in the absence of scientism, it is not clear what determines whether a given fact or

---

<sup>65</sup> Finlay (2008a, 2011); Kirchin (2010); Loeb (2010).

<sup>66</sup> Hare (1981: 86, 1999: 2); also see Lillehammer (2004: 97).

<sup>67</sup> Timmons (1999: 75); also see Nagel (1986: 144); Railton (1989, 2010); Finlay (2008a: 348).

<sup>68</sup> Brink (1984: 120).

<sup>69</sup> Nagel (1986: 144).

property is or is not an acceptable addition to our ontology. In particular, the explanatory value of a property often justifies such an addition. Third, it is not even clear that ontological issues play any (major) role in ethical and indeed metaethical theorizing at all.

A fourth problem I discuss is this. Error theories should have a theory about what to do with moral talk and thought once it is agreed that these are ‘in error’. This is not strictly speaking necessary. It is already an interesting result that *current* moral discourse is in error. But given the importance we generally attach to moral thought and talk, it would be a good thing if error theorists could tell us what to do with it after accepting their theory. The problem is that it is not clear that error theorists have so far succeeded in supplying this kind of additional theory that we can accept.

## **1.5 Conclusion and Preview**

This chapter has introduced error theory via an introduction of metaethics. It has also explored the four main problems for error theory, which, at least in conjunction, should make us suspicious of its truth. The remainder of the thesis argues that we can carve out a more plausible position in logical space for error theory by adopting different theoretical commitments than the error theories that currently exist have so far been able to do. In order to show this I start with an exposition of current error theories and argue that their attempts to solve the four main problems fail (Chapter 2). I then introduce different theoretical options for error theorists (Chapter 3) and argue that commitments different from those theories discussed in Chapter 2 get us a more plausible error theory, at least insofar as the first three problems are concerned (Chapters 4-6). I then argue that further problems for error theory other than the four main ones I have already mentioned can also be solved (Chapter 7). Finally, I argue that a hitherto highly underexplored option for what to do with moral discourse after error theory has a lot to be said in favour of it, which shows that the fourth problem just mentioned is solvable as well (Chapter 8). These results amount to progress in the debate about error theory.

## Chapter 2

# Existing Error Theories

### 2.1 Introduction

This chapter explores existing error theories, and their ways of solving the four main problems for error theorists as discussed in §1.4. It also argues that none of these attempts are very successful. This motivates the project of the remainder of the thesis; viz., to argue that a different formulation of moral error theory can do better.

I first explain Mackie's error theory. I discuss, after a brief critique of Mackie's earliest thoughts on the matter (§2.2.1), how his later work aims to settle the Non-Negotiable Commitment Claim (§2.2.2), how it argues for the Substantive Claim (§2.2.3), and what it recommends we should do with our error-ridden moral discourse (§2.2.4). The chapter proceeds with an explanation of Joyce's error theory and its Non-Negotiable Commitment Claim (§2.3.1), its Substantive Claim (§2.3.2) and its recommendation for what to do with moral discourse after error theory (§2.3.3). The chapter closes with a similar, but briefer, discussion of other error theories; namely, Olson's, Streumer's, Loeb's and Schiffer's (§2.4). A final section concludes and looks ahead (§2.5).

### 2.2 J.L. Mackie

#### 2.2.1 Mackie's Early Scepticism

Mackie first formulated his error theory in a paper from 1946. In that paper he starts by presenting an early version of the Non-Negotiable Commitment Claim:

[m]ost of us ... *judge* that actions and states are right and good, just as we judge about other matters of fact, that these judgments are either true or false, and that the qualities with which they deal exist objectively. This view, which almost everyone holds, may be crudely called "believing in morals."<sup>70</sup>

---

<sup>70</sup> Mackie (1946: 77-8; also see 80, 81).

In contrast to his later 1977 book *Ethics: inventing right and wrong*, the paper says hardly anything in defence of this claim. I think it is safe to suppose that Mackie thought that it is simply obvious and undeniable that we ‘believe in morals’. This is of course true in a sense; virtually everyone moralizes. But the question is whether a *philosophical account* of the datum that we all ‘believe in morals’ requires the postulation of the robust metaphysical picture of moral properties that Mackie proceeds to use for his moral nihilism; viz., that the “qualities with which [moral judgments] deal exist objectively.” Roughly, this is the picture on which moral properties are both wholly independent from us and yet able to affect our actions, normatively and motivationally. The problem is that this philosophical claim is neither obviously entailed by the simple observation that ‘everyone moralizes’, nor is it argued for by Mackie in this paper. Instead after making the remark cited above Mackie immediately proceeds to discuss four arguments for the claim that there are no moral values understood in the metaphysically robust way I just explained. So the first point of critique of Mackie’s early error theory is that it fails to establish the Non-Negotiable Commitment Claim.

The second point of critique is that Mackie’s four arguments for the Substantive Claim don’t work. The first two arguments resemble the famous disagreement and queerness arguments that can also be found in the 1977 book.<sup>71</sup> Argument one is an argument to the best explanation. We are better off explaining the persistence of inter- and intra-cultural disagreement about morals on the basis of the non-existence of moral facts (both parties to the dispute wrongly believe that morality exists) than on the basis of cultures or individuals within cultures differing in their epistemic access to objective moral facts (some cultures or individuals get morality right and other cultures simply lack those cultures’ or individuals’ moral sophistication). Moreover—and this is the second argument—moral facts would be very queer entities, which is further reason for not postulating them.

The third argument is that we have a good explanation of *how* and *why* it is that moral judgments feel to us as though we are describing an objective moral reality even though our “feelings [of approval and disapproval] are *all* that exists.”<sup>72</sup> The idea is that the availability of this explanation renders unnecessary an

---

<sup>71</sup> Mackie (1946: 78).

<sup>72</sup> Mackie (1946: 86, emphasis in original).



alternative, ontologically more costly explanation according to which it is the occasional latching on to moral properties that explains moral judgments' distinctive phenomenal character. Mackie's answer to the question *how* it is that we objectify our feelings is that we project our moral sentiments onto external objects and actions.<sup>73</sup> This is meant to be similar to how we project our feeling of disgust, as we get it from seeing a foul mushroom, onto the mushroom itself. And *why* do we objectify? Mackie's answer is that it is useful.<sup>74</sup> We are more inclined to refrain from killing—and hence to display behaviour that fosters stable social communities the fruits of which are to our own benefit—if we believe that killing really has the property of being morally wrong compared to when we think that all we are doing is projecting our feelings onto morally neutral actions.

Note that none of these arguments can show that we have conclusive reason to believe that error theory is true. All these arguments can show is that we have *some good* reasons to believe that error theory is true, but this is consistent with the existence of moral facts.

The fourth argument for error theory, which appears towards the end of the paper, involves an alleged absurdity that is ingrained in our ordinary use of 'obligation'. The absurdity is that obligations require both indeterminism and determinism. This argument, when successful, will be able to show that error theory is true. For no theory that entails an absurdity can be true, and any moral success theory would carry this entailment.<sup>75</sup>

Moral obligations require indeterminism because the '*ought implies can*'-doctrine is true. For it to be true that Jane ought to  $\phi$  it must be possible for Jane to  $\phi$ .<sup>76</sup> But  $\phi$ -ing is only possible for Jane if indeterminism is true: compatibilist accounts of metaphysical freedom of the will (according to which freedom is compatible with determinism) are false.<sup>77</sup> Unfortunately, obligations require determinism as well. For when we assess Jane's behaviour—and we invariably do, a

---

<sup>73</sup> Mackie (1946: 81).

<sup>74</sup> Mackie (1946: 80).

<sup>75</sup> Mackie (1946: 86).

<sup>76</sup> In this and the coming few paragraphs we read, to make sense of Mackie's writings, 'ought' as 'is obligated'.

<sup>77</sup> Mackie (1946: 87).

claim anticipating Strawson's famous 1962 essay *Freedom and Resentment*—we assess *Jane's* behaviour, not whatever forces outside of Jane's control interfered in the process leading from Jane's decision to act to her actual action.<sup>78</sup> These forces could be brain processes 'within' Jane that Jane doesn't control, or they could consist in acts of Divine intervention, or what have you. Either way, for Jane to be praiseworthy for her actions, her actions have to "flow from" her character or self "inevitably."<sup>79</sup> We now have an absurdity, for this entails that obligations require both indeterminism and determinism.

This argument has dropped out of Mackie's book. That was a good idea. The argument is not persuasive. First, it doesn't justify sinking moral discourse even when it is sound. The argument isn't powerful enough for that. As Mackie himself recognizes, moral discourse contains much more than ascriptions of obligations—it also contains for example evaluative as opposed to deontic appraisals of character traits.<sup>80</sup> Perhaps, then, moral discourse can survive the blow that its obligation-ascription judgments are all systematically false as long as its judgments involving evaluative concepts aren't. But Mackie's paper doesn't consider this. So his argument is underdetermined.

Second, Mackie's argument doesn't even seem sound. The premise that compatibilism is false is shaky.<sup>81</sup> Moreover, even if we are incompatibilists, it is not clear that for true ought statements of the form 'Jane ought to  $\phi$ ', we need to be able to praise *Jane* as something over and above the workings of her brain. This is precisely the positive proposal of *event-causal* libertarians about free will.<sup>82</sup> The fact that event-causal libertarianism is coherent means that it merits positive argumentation to rule it out. But Mackie doesn't supply this argumentation.

As Mackie's book on moral skepticism covers much the same terrain as the 1946 paper (except the argument just mentioned), and much more besides, I shall in what follows just focus on the book.

---

<sup>78</sup> Skorupski (2007: 95) notes that "moral obligation and blame are connected: if A has a moral obligation to  $\phi$  then A can be blamed for not  $\phi$ -ing."

<sup>79</sup> Mackie (1946: 88).

<sup>80</sup> Also see Kirchin (2010: 172).

<sup>81</sup> For a forceful argument for compatibilism, see Fischer and Ravizza (1998).

<sup>82</sup> Kane (1996).

### 2.2.2 Mackie on the Non-Negotiable Commitment Claim

The opening statement of Mackie's *Ethics* is that "[t]here are no objective values."<sup>83</sup> This sceptical claim should first be distinguished from other claims that might be conflated with it. It is different from *subjectivism* understood as the claim that everyone ought to do as he thinks he should.<sup>84</sup> For subjectivism so understood is a first-order moral claim, whereas Mackie's skepticism is a second-order claim. It should also be distinguished from other second-order claims. Some of these involve questions of meaning. One example is the question whether a statement such as 'this action is right' *means* 'I approve of this action', and which second-order subjectivists answer in the affirmative. Mackie makes it clear that his skepticism differs from all such semantic second-order positions because his is fundamentally an ontological second-order view.<sup>85</sup>

However, Mackie isn't *just* arguing that objective moral values are not "part of the fabric of the world."<sup>86</sup> He is also making claims about moral language—claims about what ordinary speakers of the English language mean when they utter moral statements. Mackie makes these further claims because he wants to avoid his skepticism being "trivially true."<sup>87</sup> He doesn't want it to be the case that although there are no objective moral values, no one thought there were such values anyway. In particular, his concern is that no one ever thought that moral values were objective in the *technical sense of objectivity* he has in mind; namely, the sense in which the objectivity of values is given by their existence

prior to and logically independent of all [human] activities [such as] preferring, choosing, recommending [and] condemning<sup>88</sup>

---

83 P5 (all references are to the 1977 book unless otherwise indicated). Note that Mackie, following the tradition of his time and that he is rejecting against, places (intrinsic) goodness at the normative centre stage, as opposed to notions like reasons or rationality, as is nowadays more customary (Olson 2009b: 164). Note too that Mackie says that much of what he claims about objective moral values also applies to objective *aesthetic* values, but it is not clear that it does (see Chapter 6 of this dissertation). For now I bracket the issue of skepticism about aesthetic values.

84 P17.

85 P18.

86 P15.

87 P15, 30.

88 P30; also see Joyce (2001: 16, 2011b: 157).

Call this kind of moral objectivity Mackie-Objectivity:

**Mackie-Objectivity**

Moral values are prior to and logically independent of all human activities

The claim that moral properties, to *count* as moral properties, would have to be Mackie-Objective is an extremely radical thesis. Darwall, Gibbard and Railton call this view *Ethical Platonism*.<sup>89</sup> Tresan uses the term *SuperDuper Objectivism*.<sup>90</sup> The thesis is radical because for Mackie moral properties are both existentially mind-independent and conceptually mind-independent; moral properties would exist even if no minds existed and the concept of a moral property can be adequately articulated without making reference to any mental entities.<sup>91</sup> So various kinds of constructivist, irrealist cognitivist and naturalist realist theories (insofar as these offer a different conception of moral values) are out almost before we even got started.

But moral properties are not the only items that are Mackie-Objective. Consider the stars, the sun and the moon, and the oceans, to name just a few. What sets out moral properties and facts from other Mackie-Objective properties and facts is that the former can ‘bind’ humans to perform certain actions. What this means is that moral facts and properties are ‘prescriptive,’ which Mackie understands to concern both motivational and normative force:

Moral ... values ... are a very central structural element in the fabric of the world. But it is held also that knowing them or ‘seeing’ them will not merely tell men what to do but will ensure that they do it, overruling any contrary inclinations<sup>92</sup>

---

<sup>89</sup> Darwall *et al.* (1992: 141n59).

<sup>90</sup> Tresan (2009: 369).

<sup>91</sup> Joyce (2009).

<sup>92</sup> P31-2. Scanlon (2007: 5) and Pauer-Studer (2009: 186) note that Mackie, under the influence of Hume (Mackie would write a whole treatise on Hume’s moral theory in 1980, and his 1977 book features many references to Hume), was deeply impressed by the requirement that metaethics must answer the ‘problem of moral motivation’. One explanation of why Mackie focusses on motivation when he talks about ‘prescriptivity’ is that he “seems to have confused the motivational import of [an atomic] moral belief with the normativity of a moral fact” (Copp 2010: 146; this

The first sentence in this citation reaffirms Mackie's view that moral values are objective in the technical sense that they would literally have to be part of reality itself; their mode of existence is wholly unaffected by our dealings with or our efforts to modify or interpret that reality. The second sentence explains what the prescriptivity of moral values amounts to. Moral values are normative so long as we take the relation of 'telling' someone to do something to be a normative relation. And moral values are also motivationally efficacious because they *ensure* that agents who are aware of them will do what they have to do, perhaps absent lethargy, listlessness and the like. So Mackie-Prescriptivity can be defined as follows:

**Mackie-Prescriptivity**

Moral values provide both normative and motivational direction for agents who are aware of them

What is not quite clear is whether Mackie thought that the normative or motivational force of moral properties and facts would have to be *overriding* in the following sense. Once agents are acquainted with moral facts and properties, they are automatically motivated to act in accordance with them, regardless of how strong their motivation was to do something else. Moreover, the normative force of moral facts and properties also overrides whatever else spoke in favour of the agent performing a different action. This notion of overridingness is not represented in my description of Mackie-Prescriptivity because its presence doesn't improve Mackie's argument for error theory.<sup>93</sup> If anything, it will make it harder for Mackie's error theory to be true, not easier. For the stronger the non-negotiable commitment of ordinary moral discourse is formulated, the harder it becomes to argue that moral discourse carries it. And the commitment to Mackie-Objective and Mackie-Prescriptive values is already, or so it seems, strong enough to get some kind of

---

is also Dreier's 2010 interpretation of Mackie). On Parfit's interpretation of Mackie's (1977), the argument is all about motivation (2012: 448-52). Garner (1994: 144) argues that Mackie did not distinguish clearly enough between motivational internalism and reasons internalism. For similar points see Brink (1984: 114, 114n8); Shafer-Landau (2003: 55). I address these issues in what follows.

<sup>93</sup> For arguments that non-overridability is part of our ordinary concept of morality, see McNaughton (1988: 114).

queerness argument going. Hence because I think that Mackie's error theory doesn't work—my own error theory is different from Mackie's—I want to set up my opponent's view as strongly as possible and therefore leave non-overrideability out of my account of Mackie on moral values. From now on when I write about objectivity and prescriptivity I'll have Mackie-Objectivity and Mackie-Prescriptivity in mind.

Mackie thinks that moral facts construed as both objective and prescriptive are a non-negotiable commitment of moral discourse:

the traditional moral *concepts* of the ordinary man as well as the main line of western philosophers are concepts of [objective and prescriptive] value<sup>94</sup>

This is evidence that Mackie thinks that the Non-Negotiable Commitment Claim can be settled by means of the Conceptual Entailment Claim:

**Conceptual Entailment Claim**

The mode of commitment of moral discourse to Mackie-Objective and Mackie-Prescriptive values is one of conceptual entailment<sup>95</sup>

Roughly, according to Conceptual Entailment Claim, moral statements entail statements about objective and prescriptive values because it is necessary to the correct applicability of moral concepts, as expressed by moral statements, that there exist objective and prescriptive values.<sup>96</sup> To see how this works, first consider a non-

---

<sup>94</sup> P35; emphasis mine.

<sup>95</sup> This interpretation is accepted by almost everyone in the field: Smith (1994: 64); Svavarsdóttir (2001: 145); Burgess (2010); Olson (2011a: 79n.3). However, some of Mackie's commentators write as though Mackie favoured a *presupposition* rather than an entailment approach (Burgess 1998: 543; Finlay 2008a; Copp 2010: 144). Given that presuppositions are very different from entailments (see Chapters 3-4) and given that Mackie explicitly says that it is part of our *concepts* that moral facts about objective and prescriptive, I set this interpretation aside (I suspect that the second group of commentators hasn't thought in much detail about the difference between conceptual entailment and presupposition).

<sup>96</sup> For this way of thinking about conceptual entailments see Finlay (2008a: 347); other ways of thinking about the Conceptual Entailment Claim are canvassed in Chapter 3. This is a rough characterization of the Conceptual Entailment Claim because, for instance, 'correct application conditions' is ambiguous between 'conditions of true application' and 'conditions which constrain how competent

moral example. The sentence ‘this is a vixen’ entails ‘this is a fox’ in virtue of the concept VIXEN expressed by the term ‘vixen’ in the sentence ‘this is a vixen’. For the conditions on the correct application of the concept VIXEN are two-fold: in order to use VIXEN correctly, you have to apply it to something that is both female and a fox. Take either femaleness or foxhood out of VIXEN and the *schvixens* you end up referring to simply don’t deserve the label ‘vixen’. Likewise, ‘stealing is morally wrong’ entails ‘there exist objectively prescriptive values’ in virtue of the concept MORALLY WRONG expressed by the term ‘morally wrong’ in ‘stealing is morally wrong’.<sup>97</sup> For the conditions of correct application of the concept MORALLY WRONG are two-fold: in order to use MORALLY WRONG correctly you have to apply it to something that both instantiates an objectively prescriptive value and that is of recognisable moral import.

How does Mackie aim to settle the claim that we will be using moral concepts wrongly if we don’t apply them to things that both instantiate an objectively prescriptive value and that is of recognisable moral import? Take recognisable moral import first. Mackie thinks that the *subject-matter* of moral judgments concerns social coordination—the real function of moral systems is social coordination.<sup>98</sup> Let us simply accept that this is true in line with my earlier proposal to run with the idea that we can recognise the moral when we see it (see n3 above). So let’s simply assume that when social usefulness is at stake our discussion becomes morally relevant. The beef in Mackie’s account is in his claim that it is part of moral concepts’ correct application conditions that they have to be applied to things that are objectively prescriptive. He has two arguments for that claim. One argument focuses on how philosophers use moral concepts (capturing, at least to a large enough extent, claims Mackie, what the folk are after).<sup>99</sup> The other argument focuses on how the folk use moral concepts.

---

concept-users apply the concepts’ and perhaps other readings as well. These details are unimportant in the present context and will be discussed in the next chapter.

<sup>97</sup> We can take, with Parfit (2012 Vol2: 333), ‘stealing is morally wrong’ to be equivalent to ‘stealing morally ought not to be done’, which in my experience to some ears more clearly licences the claim that moral values are objectively prescriptive.

<sup>98</sup> Chapter 5.

<sup>99</sup> “I think that Kant is struggling to bring out something that is latent in ordinary moral thought [he is] not merely constructing a philosophical fantasy” (P60).

Let's first look at the argument focusing on how philosophers use moral concepts. Mackie tries to show that the best interpretation of their work reveals that they think that using moral concepts entails describing things that are objectively prescriptive. Mackie looks at Kant, Aristotle, Hutcheson, Price, Plato and Sidgwick. For instance, according to Clark there are:

eternal and necessary differences of things that make it *fit and reasonable* for creatures ... to act [in a certain way]<sup>100</sup>

Moral values are Mackie-Objective because they are 'eternal and necessary' (and things that are 'eternal and necessary', the assumption goes, are prior to and logically independent of all human activities). And moral values are Mackie-Prescriptive because they make it 'fit and reasonable' for agents to act in a certain way (and the relation of making it 'fit and reasonable' for agents to act in a certain way is a normative relation).<sup>101</sup>

What to make of this argument? One thing we can wonder about is whether the argument's interpretation of these philosophers' work is correct.<sup>102</sup> The problem with Mackie's remarks is that they are too quick to count as serious exegesis. We can also ask why the work of those philosophers who have proposed a less ontologically demanding understanding of the nature of moral facts and properties can be ignored, as Mackie does. Finally we can ask for a justification of the claim that these philosophers are trying to put into words what the folk are committed to when they participate in moral discourse. Mackie doesn't supply this justification. Hence this argument of Mackie's is shaky to say the least. Perhaps then Mackie's

---

<sup>100</sup> P31.

<sup>101</sup> In fact this quote reveals one problem with my interpretation of Mackie-Prescriptivity. For according to my interpretation Mackie-prescriptivity involves both normativity and motivation but this quote only involves normativity. In light of numerous other remarks elsewhere in the text that do take prescriptivity to involve both normativity and motivation, I put this down to a slip of the pen on Mackie's part.

<sup>102</sup> Take Kant. Davidson (2004: 40) interprets Kant as a realist in Mackie's sense. Rauscher (2002) dissents with an interpretation of Kant according to which he is committed to a stark form of moral anti-realism. Similar questions arise for the other philosophers.



claim that the Conceptual Entailment Claim can be settled by looking at the folk directly can bring solace.

Mackie draws attention to the following pattern of reactions that the folk display to a perceived need to abandon their belief in moral values, such as it may assail them when they learn that error theory is true. This abandonment causes “at least temporarily, a decay of subjective concern and sense of purpose.”<sup>103</sup> The fact that this happens, argues Mackie, is evidence that people objectify their purposes; imbuing them with a fictional external authority as captured by their values’ objective prescriptivity. After all, if we hadn’t thought that moral facts were ‘real’ in the robust sort of way that Mackie argues they would have to be then this decay of subjective concern and loss of purpose presumably wouldn’t occur. You presumably only feel ‘lost’ when you learn that there are no ‘real, objective values’; not when you learn that the values you’ve been making up don’t exist. Mackie wants to infer from this that “it is not going too far to say that this assumption [of moral values’ objectivity and prescriptivity] has been incorporated in the basic, conventional, meanings of moral terms.”<sup>104</sup>

This can be doubted in three ways. First, we can ask why we would be justified in thinking that this explanation of the folks’ behaviour is better than other explanations. Consider the following, rival explanation. The folk merely have a *false belief* about the referents of their moral concepts.<sup>105</sup> They think, falsely, that morality is about objectively prescriptive values whereas in fact morality is about actions for which there exists no objective prescriptivity to perform them. We can also explain the datum that the folk feel lost when they learn about error theory by reference to this false belief rather than by reference to the claim that is part of moral concepts that there must be objective prescriptivity. The rival explanation is that the folk’s big myth around moral values must be re-written; that they have been wrong about what moral values are, which are near and dear to them. Re-writing this myth will be difficult and painful, and learning to live with the truth might cause the folk to have

---

<sup>103</sup> P34. Also see Hare (1999: 2).

<sup>104</sup> P35, for “[a]ny analysis of the meanings of moral terms which omits this claim to objective, intrinsic, prescriptivity is to that extent incomplete; and this is true of any non-cognitive analysis, any naturalist one, and any combination of the two” (P33).

<sup>105</sup> Finlay (2008a); Tresan (2009: 369); also see Gill (1996).

the reaction to error theory described earlier. The problem for Mackie is that he hasn't considered this rival explanation.

Second, supposing that the earlier question can be answered, we can still ask: why do we need to translate the 'robustness' of moral values—which we (by hypothesis) need to postulate to explain the reactions of the folk to learning about these values' non-existence—in terms of Mackie-Objectivity and Mackie-Prescriptivity? There are other notions of 'robustness', captured for instance by *inter-subjectivity* rather than objectivity.<sup>106</sup> These notions might do the explanatory job equally well here. They too might allow us to explain the alleged datum about folk reactions. But, again, Mackie doesn't consider this option.

Third, is it really true that if we learn that it is not morally right, laudable or obligatory to help others who suffer because nothing is morally anything we will experience a decay of subjective concern and sense of purpose? Mightn't we continue to care about the *suffering* of people all the same, without decay of subjective concern, simply because it is the *suffering* we care about rather than whatever further objective-cum-prescriptive moral property this suffering is said to instantiate? If that is so then again Mackie's argument will be compromised. For then the datum that Mackie argues requires for its explanation the postulation of objectively prescriptive moral properties—viz., that people will feel lost after hearing about error theory in the way that he says they will—won't even be a datum.

The combined force of these objections to Mackie's argument entails that his attempt to establish the Conceptual Entailment Claim by looking at folk behaviour fails just like his attempt to establish the Conceptual Entailment Claim by means of interpreting philosophical texts about morality.

To sum up our discussion so far, we have discussed what Mackie means, exactly, when he denies that there exist moral values: moral values would have to be both Mackie-Objective and Mackie-Prescriptive. We have also explained that Mackie couples this claim with a second claim: the folk as well as the moral philosophers believe that morality and moral discourse are non-negotiably committed to moral values thus understood. Finally we saw that Mackie's arguments

---

<sup>106</sup> Burgess' (2007) error theory goes in terms of inter-subjectivity rather than (Mackie-)objectivity.

for the second claim don't even come close to settling the Non-Negotiable Commitment Claim.

Perhaps, then, we have been looking in the wrong direction. I have so far presented the standard interpretation of Mackie.<sup>107</sup> But there is another interpretation according to which what is queer are not moral facts and properties themselves but moral beliefs and their alleged motivating power (motivational internalism) or moral beliefs and their alleged power to provide reasons for action (reasons internalism).<sup>108</sup> Let us therefore look at these interpretations of Mackie to see whether they can help his case.

First consider motivational internalism, and consider the Anti-Humean Theory of Motivation:

At least some beliefs, perhaps in addition to desires, are motivationally efficacious mental states<sup>109</sup>

---

<sup>107</sup> Brink (1984: 113); Hampton (1998: 21-6); Joyce (2001: 16-29); Dreier (2010: 76); Garner (1990).

<sup>108</sup> Copp writes: "Mackie thinks that the normativity of goodness would consist at least in part in the fact that, necessarily, the *belief* that truth-telling is good would provide anyone with a motive to be truthful" (2010: 145, emphasis added). Brink writes: "In claiming that moral facts would have to be objectively prescriptive, Mackie is claiming that moral realism requires the truth of *internalism*. Internalism is the *a priori* thesis that the recognition of moral facts itself either necessarily motivates or necessarily provides reasons for action." (1984: 113). A third interpretation focuses on Mackie's theory of practical reason rather than moral metaphysics (as the standard interpretation has it) or moral beliefs and various kinds of internalisms (as Copp and Brink have it). This third interpretation is defended by Schroeder (2007: 119 n23); Philips (2010: 87, 95); Joyce (2001: 42, 77, 2005); Olson (2009b: 177); Robertson (2010: 435). However, this third interpretation can be dismissed on the ground that Mackie writes that talk about imperatives expressing categorical reasons makes the issue about the non-existence of moral values not existing "clearer" (P27). I take it that the most natural interpretation of this is that moral discourse's commitment to moral *properties* is more fundamental than the commitment to practical reason. I won't therefore discuss this third option in the text. In the text I do argue that Copp's and Brink's interpretation doesn't have a lot going for it.

<sup>109</sup> The Anti-Humean Theory of Motivation negates the Humean Theory of Motivation according to which beliefs by themselves never motivate and can only motivate, when they do, if conjoined with motivationally efficacious desires (Smith 1994: Ch.4).

If the Anti-Humean Theory of Motivation is true then our sense that morality is prescriptive can be explained without postulating moral facts outside of the agent that have a motivational pull on the agent. For if the Anti-Humean Theory of Motivation is true then morality can motivate because our *beliefs* about morality can motivate. In that case Mackie could take it as his Non-Negotiable Commitment Claim that morality requires the truth of the Anti-Humean Theory of Motivation and he could take it as his Substantive Claim that the Anti-Humean Theory of Motivation is false. Perhaps beliefs that are intrinsically motivationally efficacious are queer.

There are various problems with this suggestion. One problem is that although Mackie accepts the Humean Theory of Motivation in the *Ethics*, in his later book on Hume's ethical theory he rejects it in favour of the Anti-Humean Theory of Motivation.<sup>110</sup> Another problem is that the Anti-Humean Theory of Motivation doesn't seem to be able to do enough work for Mackie. Mackie wasn't just worried about morality's motivational pull; he was worried about its *guaranteed* motivational pull (which is not to imply that Mackie was worried about morality's *overriding* motivational pull):

values [have a] power, when known, *automatically* to influence the will<sup>111</sup>

But the Anti-Humean Theory of Motivation is consistent with values *not* automatically influencing the will. The Anti-Humean Theory of Motivation only implies that beliefs *can* motivate, not that they are *guaranteed* to motivate.

One could try to remedy this by accepting a slightly modified version of the Anti-Humean Theory of Motivation:

At least some beliefs, perhaps in addition to desires, are *necessarily* motivationally efficacious mental states

However, this spurs the worry that the natural conclusion of Mackie-style arguments against morality would not be that morality is flawed but that moral judgments are

---

<sup>110</sup> Mackie (1980: 53, 54, 141-2).

<sup>111</sup> P40; emphasis added.

not beliefs, which sits ill with Mackie's rejection of non-cognitivism. After all, if it is essential to moral judgments that they are beliefs and necessarily motivate, but in fact cannot motivate, then, as one natural conclusion has it, moral judgments are not beliefs.<sup>112</sup>

But most importantly, this alternative interpretation of Mackie on the Non-Negotiable Commitment Claim also sits ill with what Mackie does and doesn't say in his *Ethics*. He nowhere says that beliefs or judgments motivate, but he does say that *knowledge by acquaintance* of moral facts guarantees motivation (see the last quote from *Ethics* above). Thus it seems that for Mackie, with his theoretical commitment to the truth of cognitivism, we would have to postulate objectively prescriptive *properties* to account for morality's objective prescriptivity, as my earlier interpretation commits him to. So this alternative account of morality's poisoned commitment according to which what is queer are moral beliefs rather than moral properties doesn't, as presented, get us a working moral error theory either.

The problems I just rehearsed for the motivational internalism-interpretation of morality's alleged poisoned thesis also apply, *mutatis mutandis*, for the reasons-internalism interpretation of morality's alleged poisoned thesis. According to this proposal, what Mackie thinks moral discourse is non-negotiably committed to are beliefs that provide reasons for action and what Mackie finds queer is that such beliefs exist. The most important problem with this interpretation, again, is that Mackie nowhere says that it is *beliefs* that supply moral reasons whereas he does say that knowledge by acquaintance of moral facts provides morality's required normativity. So the reasons-internalism interpretation too should be ignored.

Overall then, Mackie's claim that moral discourse carries a non-negotiable commitment is in serious trouble. Internalism interpretations sit ill with Mackie's text. The mind-independent, Mackie-Objective and Mackie-Prescriptive moral properties-interpretation sits much better with the text, but Mackie's arguments for the claim that morality couldn't be *morality* without the postulation of moral properties thus understood are highly inconclusive.

---

<sup>112</sup> Shepski (2008: 373); Olson (forthcoming-b).

However, from now on let's assume, *arguendo*, that Mackie has a case for his 'conceptual' claim. Let's turn to his arguments for the Substantive Claim that there are no objectively prescriptive moral facts and properties.

### 2.2.3 Mackie's Substantive Claim

Mackie has three arguments for the Substantive Claim that there is no objective prescriptivity. I discuss them in turn.

The first is the argument from queerness. Moral values would have to be part of the universe to count as objective in Mackie's sense because in that sense of objectivity things are objective if and only if they are independent of and logically prior to all human activity. They would also have to be prescriptive in the sense of supplying both normative and motivational force. Now according to Mackie our world contains *subjective* prescriptivity, such as values for sheepdog trials and the rules of etiquette (to be discussed later in this section). Our world also contains objective *non-prescriptive* things, such as the stars and the oceans. But our world doesn't contain objective prescriptivity. Such things would be too 'queer' to exist. We can formalise the argument as follows:

#### Queerness Argument

- P1     Morality's non-negotiable commitment: moral motivation stems from the motivational pull of mind-independent moral properties exerted on the human psyche and moral normativity is provided by the same mind-independent moral properties
- P2     Any fact that exerts motivational pull on the human psyche (i.e., any fact such that knowledge by acquaintance of that fact guarantees that the knower is motivated) is queer and any fact that tells agents what to do that is not suitably connected to those agents and their desires, plans and social roles *et cetera* is queer
- C1     Hence, moral facts are queer [From P1, P2]
- P3     We can safely assume that queer things don't exist
- C2     We can safely assume that moral facts don't exist [From C1, P3]<sup>113</sup>

It is not clear that this argument works. Most importantly, we need to know what the queerness alluded to involves. What, exactly, does it take to be queer, and on the basis of what is the inference in P3 from queerness to non-existence justified? We can distinguish, with Shepski, between three ways in which properties can be

---

<sup>113</sup> Cf. Olson (forthcoming-b); also see Burgess (1998: 543); Skorupski (2007: 92).

queer: moral properties can be ontologically proliferate, *sui generis* and mysterious (i.e., ‘incomprehensible’ or ‘unexplainable/ineffable’).<sup>114</sup> But as Shepski argues, none of these interpretations of queerness gets us a working argument *from queerness*.

Start with being ontologically proliferate. If postulating an entity or property expands our ontology then that by itself doesn’t make that entity or property queer. Shepski gives the example of Pegasus; there is nothing ‘queer’ about a horse with wings, it is just that such entities do not exist.<sup>115</sup> Adding flying horses to our ontology does not mean adding ‘queer’ things to our ontology. Moreover, even if we assume that an expansion of our ontology is *prima facie* queer, we often justify it on the following ground. The expansion is required for a good philosophical account of some phenomenon, such that the overall account that includes the expansion is better (explains more, *et cetera*) than an account that doesn’t make the expansion. It is not clear that this phenomenon couldn’t hold for morality.

The same holds for expansions of our ontology with *sui generis* and mysterious properties. Take mysteriousness first. The fact that a property is ‘mysterious’ doesn’t make that property queer. The nature of our failure to comprehend something is important here. If we model our failure to comprehend moral values on our failure to comprehend ‘round squares’ then we don’t have an argument *from queerness*. For in that case we would have an argument *from the meaningless of the language* we are using to try to pick something out in the world.<sup>116</sup> Another model for ‘failure to understand’ would be taking it as something

---

<sup>114</sup> Shepski (2008). Heumer (2005: 200) suggests a fourth way in which properties can be queer; viz., by being ‘counter-intuitive’, but from his discussion it becomes clear that he has Skepski’s ‘mysteriousness’ in mind.

<sup>115</sup> Shepski (2008: 375).

<sup>116</sup> Suggestions that Mackie employs this argument have been introduced into the debate by Williams (1985b: 203-4). Others have picked up. Smith writes that for Mackie, “[t]he problem with ... [moral properties] isn’t that there aren’t any such things as a matter of contingent fact. The problem is that we can literally make no sense of them” (2010: 119; also see 1993: 237). Copp writes that according to Mackie, “There could not be facts that, as a matter merely of the way they are in themselves, necessarily motivate people who are aware of them.” (2010: 145). Nolan et al. (2005: 325-6) write: “nothing could even possibly have the sort of objective prescriptivity Mackie describes ... [so that positive moral claims are false] of semantic or logical necessity.” Also see Lillehammer (2004: 96); Kalderon (2005:

like ‘beyond human understanding’. However, although inferences from our failure to *conceive* of something to that thing’s non-existence are fairly standard in philosophy, inferences from ‘failing to be understood’ to ‘non-existent’ aren’t recognized. For an inference from *our* failure to understand something to that thing’s queerness and non-existence is clearly unwarranted. A third model for ‘incomprehensible’ would be ‘currently unexplained’. But things aren’t queer because we currently can’t explain them. The objects and properties that quantum science postulates aren’t queer because we currently do not fully understand them. For all these cases, moreover, it also holds that if the postulation of ‘mysterious’ properties is nonetheless required for a good philosophical account of some phenomenon, then we aren’t justified to infer from the alleged *queerness* of the property that it doesn’t exist. On the contrary, adding some queerness to our overall account of the phenomenon might be a price worth paying. Again it is not clear that this phenomenon couldn’t hold for morality.

Consider, then, the interpretation of queerness as *sui generis*. Can being different from everything else in the universe make a property queer? Again off the cuff it doesn’t seem like it does. In a way, everything is different from everything else and some things are *very* different from all the other things. The property of being alive, perhaps, is a case in point.<sup>117</sup> Is the property of being alive therefore queer and, because of that, non-existent? It seems that the answers would have to be ‘no’ and ‘no’. Moreover, again, if the postulation of *sui generis* properties is nonetheless required for a good philosophical account of some phenomenon, then we aren’t justified to infer from the alleged *queerness* of the property that it doesn’t exist. And again we seem to have no good reason not to extend this line of reasoning to morality.

Thus far I have mentioned two problems with Mackie’s queerness argument. First, it is not clear that at least one of the three specifications of queerness is really a specification of *queerness* (as opposed to something else). Second, it is not clear that even if one of them is it follows that the queer property doesn’t exist. For there can

---

101); Loeb (2008); Olson (2011c: 34). I explain why I don’t think that Mackie put forward this argument for error theory shortly.

<sup>117</sup> Also consider: “The world is a queer place. I find neutrinos, aardvarks, infinite sequences of objects, and (most pertinently) impressionist paintings peculiar kinds of entities” (Platts 1980: 72); also see Enoch (2011a: 135).



be good reasons to postulate queer properties if they are explanatorily useful. Mackie, I argue, hasn't done enough to show that in fact we can get a working specification of queerness and that in fact we can show that being queer means not existing.

So how does Mackie attempt to make Queerness Argument sound? Shepski argues that Mackie takes the *sui generis* interpretation of queerness—queerness as an objectionable form of uniqueness—and cites Garner:

Moral facts are not just unusual ... they are unusual in an unusual way<sup>118</sup>

The question remains, however, whether being unusual in an unusual way is a good reason to be regarded as non-existent. This is not at all clear. Maybe in fact being alive is not just unusual but unusually unusual. But most certainly there are things that are alive. And at the risk of sounding like a broken record I add that expanding our ontologies with unusually unusual properties can be acceptable on theoretical grounds and that there is no reason to think that moral properties, even though they might *prima facie* be unusually unusual, couldn't likewise be accepted into our ontology.

Some of the arguments we just discussed in passing may well be excellent arguments for the Substantive Claim, but they aren't arguments *from queerness*. If moral properties have mutually inconsistent application conditions (recall the argument from the meaninglessness of moral language) then there are no moral properties because there couldn't *be* moral properties. But I don't think that Mackie put forward this argument. His argument isn't that moral properties are impossible; it is that they are possible but *queer*. After all, had queerness meant 'impossible' in Mackie's mouth then the arguments from disagreement and projectivism (which Mackie did put forward after the queerness argument and which will be discussed shortly) wouldn't have been necessary. But Mackie nowhere says that they are strictly speaking superfluous, so queerness does not mean 'impossible' for Mackie. It would also be a misuse of the word 'queerness' to have it mean 'impossible'.

---

<sup>118</sup> Garner (1990: 143).

Perhaps then we should focus instead on the queerness of our epistemic access to moral facts to get an argument from queerness going. Here the argument is that if objectively prescriptive moral values existed then our epistemic access to them would require “a special sort of intuition.”<sup>119</sup> The postulation of this special faculty of moral knowledge is required on the following ground: you need a special mental faculty for detecting moral properties as moral properties are ontologically *sui generis*.<sup>120</sup> However, argues Mackie, the postulation of this faculty is “lame”.<sup>121</sup> This argument faces an analogous worry to the previous argument; viz., that of securing a distinctive kind of *queerness* as the main driving force of the argument. Moreover, it seems that explanatory redundancy is the main driving force behind the argument, not queerness. For, says Mackie, postulating a special faculty for moral perception is meant to be ‘lame’—not ‘queer’. Furthermore, we can once more observe that believing in queer facts can be justified on the basis of a well-worked out theory that has a lot of explanatory power and other epistemic virtues. Since the only difference between the two types of queerness arguments is that one focuses on the queerness of moral facts directly whereas the other focuses on our epistemic access to these facts, it seems that this argument, too, by parity of reasoning, is unlikely to be successful.

Mackie has a third and final queerness argument. It focusses on the *supervenience relation* that holds between natural (non-moral) facts and properties on the one hand and moral facts and properties on the other hand. It will be helpful to quote the relevant passage in full:

Another way of bringing out this queerness is to ask, about anything that is supposed to have some objective moral quality, how this is linked with its natural features. What is the connection between the natural fact that an action is a piece of deliberate cruelty ... and the moral fact that it is wrong? It cannot be an entailment, a logical or semantic necessity. Yet it is not merely that the two features occur together. The wrongness must somehow be ‘consequential’ or ‘supervenient’; it is wrong because it is a piece of deliberate cruelty. But just what *in the world* is signified by this ‘because’?<sup>122</sup>

---

<sup>119</sup> P39.

<sup>120</sup> Also see Brink (1984: 123).

<sup>121</sup> P39.

<sup>122</sup> P35.

Shepski finds in this passage evidence for the claim that Mackie's doesn't find semantic or logical necessitation (supervenience is a necessary relation) queer because these relations are explicable and what is queer about moral supervenience is that it is not explicable (rather than unusually unusual, as moral facts and properties and our ways of gaining epistemic access to these are meant to be).<sup>123</sup> He then points out that Mackie may well be right that *semantic* necessitation is explicable. Consider a creature's being an ANIMAL supervening on its being a CAT, which we can explain by reference to the definition of CAT which includes, at least, that cats are ANIMALS. However, logical necessitation isn't explicable in that way. For in the case of logical necessitation all we can do in trying to 'explain' why for instance the validity of an argument supervenes on the argument's logical form is multiply examples.<sup>124</sup> But that isn't *explaining*. Hence, argues Shepski, it either follows from the queerness argument focussed on moral supervenience that logical necessitation too would be queer and thus non-existent or else that moral supervenience isn't queer even though it isn't explicable. After all, for moral supervenience too it seems that we can only multiply examples.

This reading of Shepski's may have its merits and may well trouble this queerness argument. But I think it leaves out the important point that Mackie uses a too coarse grained notion of supervenience. It is nowadays customary to conceive of some supervenience relations as relations of *mere* modal covariance ('necessary coextension') and thus *without* any reference to ontological dependence or something being 'consequential' on something else. Often "when someone asserts that A supervenes on B, she also wants to say that A-properties ontologically depend upon B-properties ... However ... this goes beyond the minimum required for supervenience."<sup>125</sup> Relations of necessary co-variance are probably ontologically

---

<sup>123</sup> Shepski (2008: 383-4).

<sup>124</sup> Carroll (1895) comes to mind here.

<sup>125</sup> McLaughlin and Bennett (2010: §3.5, also see §3.8). Examples of philosophers who accept that ontological dependence is part of moral supervenience include the following. Shafer-Landau talks about the non-moral features of a situation *fixing* its moral status and writes that "moral facts necessarily covary with descriptive ones because moral properties are always realized exclusively by descriptive ones" (2003: 77). Brink writes: "Supervenience implies that no change can occur in the supervening property without a change occurring in the base property, but it also asserts a claim of ontological dependence" (1984: 119). Also see Moore (1942: 588); Olson (2012: 195).

innocent whereas conceptions of the supervenience relation that go beyond this need not be.<sup>126</sup> The problem with Mackie's argument is that he simply *assumes* that the *moral* supervenience relation renders moral properties ontologically or explanatorily *dependent* on the natural (or non-moral) properties. Thus Mackie's argument is underdetermined. He hasn't done enough to argue that the relation between moral and natural features of an action must be of the non-innocent kind. Furthermore, even if Mackie gets the non-innocent version of supervenience, we still face the questions whether we can make sense of a separate argument from *queerness* and why we would be justified to move from queerness to non-existence in the face of known moves in philosophy and science where queer properties are accepted in people's ontologies because of their explanatory value.

This ends my discussion of Mackie queerness argument, which is one of three arguments Mackie offers for the Substantive Claim.<sup>127</sup> I conclude that the queerness argument wasn't successful. In what follows I discuss the remaining two arguments that Mackie gives for his Substantive Claim. After that I explain what follows from the truth of Substantive Claim in terms of the truth-value of moral statements. I close by explaining what follows from the truth of Substantive Claim for the tenability of value judgments that do not entail objective prescriptivity.

The second argument for the Substantive Claim that there is no objective prescriptivity that Mackie puts forward is the argument from disagreement.<sup>128</sup> Mackie makes the empirical observation that there is an enormous amount of variation of moral values across different cultures and times but also within cultures and that moral disagreement often seems intractable.<sup>129</sup> He then argues that the *best explanation* of these data is that there are no moral facts and properties. This

---

<sup>126</sup> Kripke (1972: 153-4).

<sup>127</sup> Mackie in fact says that he has five arguments (P49) but I have contracted his argument from the queerness of moral facts, the queerness of the supervenience relation between moral facts and natural facts, and the queerness of our epistemic access to moral facts into one argument, which I call 'the queerness argument'.

<sup>128</sup> I borrow this label from Brink (1984: 112). Mackie uses the label 'argument from relativity', but this is misleading. Mackie's argument isn't 'from' relativity to some other conclusion; instead, Mackie's argument is one 'from' *disagreement* to relativity (which is meant to be inconsistent with the objectivity of moral facts and properties).

<sup>129</sup> P36-8.

explanation is meant to be better than the rival explanation that there *are* such facts and properties but that some cultures or individuals have better epistemic access to them than others. Because of this, and other things being equal, we shouldn't postulate moral properties.<sup>130</sup>

This argument can be challenged on two fronts. First, there may well be not as much relativity in moral opinions held across cultures as Mackie suggests.<sup>131</sup> However, I am prepared to give Mackie the benefit of the doubt here. I focus on my second worry, which is that we may be able to come up with a philosophical account of moral disagreement that *explains* why some cultures have better epistemic access to moral truth than others. One place to start is with the conceptually mediated nature of moral observation. Suppose moral observation is conceptually mediated in the sense that it takes taught conceptual sophistication to be able see the instantiation of moral properties. Just like I can't see that Fred told a joke without possessing a (probably fairly complicated) conceptual sophistication about social circumstances and humour, I won't be able to perceive that torturing innocent babies just for fun is wrong without possessing the required conceptual sophistication for that perception.<sup>132</sup> Suppose further that such conceptual sophistication is at least in part dependent on circumstances largely beyond the agent's control. Perhaps it depends on the education she receives. In that case, given that some cultures differ in their education, we have at least the beginnings of an explanation of such cultures' deprived epistemic access to moral truth. The explanation is that these cultures are in relevantly different circumstances.<sup>133</sup> As Mackie doesn't consider these kinds of proposals in any detail his argument fails to convince that the best explanation of said data favours moral nihilism.

Mackie's third and final argument for moral error theory is that it can explain where our false beliefs in the objectivity of moral values come from. This again would make the postulation of moral properties as the things that our beliefs latch

---

<sup>130</sup> Also see Lillehammer (2004: 97).

<sup>131</sup> Parfit (2012 Vol1: 145).

<sup>132</sup> McGrath (2004: 221); also see Denham (2000).

<sup>133</sup> See Brink (1984: 116-8) for an alternative explanation based on a coherentist moral epistemology.

onto explanatorily superfluous. Just like in his 1946 paper, Mackie aims to clinch the argument by means of defending:

**Moral Projectivism**

Moral sensations, caused by our own mental states such as wants and desires are taken to be objective, mind-independent features of reality<sup>134</sup>

Belief in the objectivity of moral values can be explained by reference to the propensity of our minds to mistake feelings that arise from within ourselves with evidence for the external world having some kind of property.

Having said this however, exactly how Moral Projectivism and moral error theory are related in Mackie's *Ethics* is subject to controversy.<sup>135</sup> Here are three possibilities. (I) By the time Mackie has discussed his arguments from queerness and disagreement he needs a tie-breaker because believing in morals is still permitted at that point. The tie-breaker would be the argument from moral projectivism. Joyce believes that the following quote is evidence for this interpretation. Immediately after his discussion of the queerness and disagreement arguments Mackie writes:

Considerations such as these suggest that it is in the end less paradoxical to reject than to retain the common-sense belief in the objectivity of moral values, provided that we can explain how this belief, if it is false, has become established and is so resistant to criticism<sup>136</sup>

The word 'provided' is important here: we can read this passage as saying that we aren't yet justified in embracing moral scepticism. We first have to explain where our moral beliefs come from and why they feel to us the way they do before we can become moral sceptics. (II) After the queerness and disagreement arguments the error theory is established, and Projectivism solely functions to explain where our error comes from for the curious reader. (III) Mackie believes in *epistemological*

---

<sup>134</sup> P42-6. As Garner puts it: "[t]he error pointed to in [Mackie's error theory] is the error we make when we take our projections for independent features of what we are evaluating, or see our requirements, or the requirements of others, as lodged in reality itself" (1994: 139).

<sup>135</sup> Joyce (2010a: 44-5).

<sup>136</sup> P42.

*conservatism*; the view that, roughly, the beliefs one brings to the table before one starts doing any sort of inquiry, including metaethical inquiry, carry a certain level of epistemic warrant that is difficult to overturn even in the face of arguments to the contrary (such as the queerness and disagreement arguments). Interpretation (III) therefore has it that Mackie musters the truth of Moral Projectivism to overcome the hurdle of epistemological conservatism.

Joyce accepts reading (III) and finds evidence for Mackie's acceptance of epistemological conservatism in the following passage:

[moral skepticism] goes against assumptions ingrained in our thought and ... conflicts with ... common sense [therefore] it needs very solid support<sup>137</sup>

On this interpretation, Mackie argues that a genealogical explanation of moral beliefs that doesn't make reference to the objects of those beliefs undermines whatever *prima facie* justification these intuitions might otherwise have had. Therefore, even though the principle of epistemological conservatism is true, according to Mackie, and threatens error theory, error theory can still be accepted.

My own view is that the first and third interpretations make good sense of our evaluations of the queerness and disagreement arguments as presented in this sub-section. For we saw that neither argument is conclusive, so further evidence for error theory is required. However, I also find the third interpretation's envisaged restriction on the role that Moral Projectivism can play in the dialectic unnecessarily restrictive. It may very well be that Moral Projectivism can provide *direct* support for moral error theory. So far we haven't seen any evidence that its only role could be to undermine the *prima facie* justification of our earlier beliefs in morality, and hence that its only role could be to counteract epistemological conservatism.

Joyce however argues that Moral Projectivism alone cannot be used as a direct argument for moral nihilism and so can only be used to undermine epistemological conservatism and its consequences.<sup>138</sup> He first distinguishes between two kinds of projectivism. A 'minimal' projectivism holds just that the experience of moral wrongness is immediately caused not by a perception of a moral property but

---

137 P35.

138 Joyce (2010a: 47).

by a moral emotion. Minimal projectivism is *metaethically neutral* in two senses. First, it is silent on the cognitivism/non-cognitivism issue: the thesis that moral experience is caused by a moral emotion is consistent with the thesis that moral judgments are non-cognitive attitudes but also with the thesis that they are beliefs. Second, minimal projectivism leaves it open that there are moral properties.<sup>139</sup> After all, the thesis that our moral experiences are immediately caused by our own emotions rather than by moral properties is consistent with the existence of moral properties. A more robust version of Moral Projectivism could solve this problem by adding to Minimal Projectivism that moral properties are metaphysically queer. If the queerness argument would work then the moral realist, untouched by minimal projectivism, is sunk (however the non-cognitivist may still be in the game). But this also means that queerness is doing all the work and thus that the projection mechanism is argumentatively idle. And, anyway, queerness arguments don't work. All of this seems right to me, and so I conclude that arguments for the Substantive Claim based on Moral Projectivism won't work.

This gets me my conclusion: although Mackie's arguments for his Substantive Claim that there are no Mackie-Objective and Mackie-Prescriptive properties are inventive and influential, they also stand in need of further support. They do not, not even in conjunction, make the case for moral skepticism sufficiently compelling.

Suppose however that Mackie's arguments for the truth of Substantive Claim are persuasive. We are justified in believing that there are no moral values. What does this mean for the status of moral statements understood as attempting to describe a moral reality that consists of Mackie-Objective and Mackie-Prescriptivity moral values? Mackie believes that his error theory makes all first-order moral judgments *false* (rather than truth-valueless, or neither true nor false, or what have you). However, moral judgments are not *necessarily* false. Mackie didn't think, some of his readers' interpretations notwithstanding, that the instantiation conditions for moral properties are somehow mutually exclusive or jointly incoherent.

I will now explain why there nonetheless are, according to Mackie, *subjectively* prescriptive values consistently with a denial of objectively prescriptive

---

<sup>139</sup> Also see Brink (1984: 111).



values. This is important for Mackie because in the absence of an argument for subjective prescriptivity his error theory would be stronger than he wants it to be. For his error theory of objectively prescriptive values will then entail error theory of normativity as such. As it turns out, however, Mackie cannot avoid this collapse of his local error theory of morality's objective prescriptivity to everything that is normative. This means that Mackie unwittingly accepts the first horn of the Formulation Dilemma of §1.4; viz., that we should be broad, meta-normative sceptics. But this horn, we said, is very implausible. This is further evidence for the implausibility of Mackie's version of moral error theory.

Mackie claims that "'X is good'-claims are always of the form 'X satisfies the requirements in question'".<sup>140</sup> Take sheepdog trials.<sup>141</sup> There are various standards that determine how well sheepdogs do in such trials. The standards are widely agreed upon by experts and they bear a close connection to the work that sheepdogs are kept to do. Good or valuable sheepdogs are those that measure up to the standards and

[g]iven any sufficiently determinate standards, it will be an objective matter, a matter of truth and falsehood, how well any particular specimen measures up to those standards<sup>142</sup>

But, importantly, the standards are based on our "choosing or deciding to think in a certain way."<sup>143</sup> So they aren't *Mackie-Objective* and prescriptive. This saves them from being 'queer', as only the combination of Mackie-Objectivity and Mackie-Prescriptivity is queer.

---

<sup>140</sup> P55. Ziff concurs: the best analysis of good was that "'good' in English means 'answering to certain interests,' the interests in question being supplied by the context in which something is said to be good (1963). In a similar vein Rawls write that "A is a good X if and only if A has the properties (to a higher degree than the average or standard F) which it is rational to want in an X, given what X's are used for, or expected to do and the like (whichever is appropriate)." (From: Scanlon 2011: 445).

<sup>141</sup> P26; also see Evans and Shah (2012: 86).

<sup>142</sup> P26.

<sup>143</sup> P16.

The same holds for the institution of etiquette.<sup>144</sup> Again it is objectively true that you have to eat peas with a fork, whether you want to or not. Nevertheless, still, these rules aren't *Mackie-Objective* because they are again based on our 'choosing or deciding to think in a certain way'. No one thinks that the rules of etiquette are set in stone as much as moral rules are. Whether eating peas with a fork or a spoon is the thing to do depends on what we (collectively) decide; whether you can kill has no such relation to our decisions. This again is what makes the rules of etiquette metaphysically respectable; only the combination of Mackie-Prescriptivity (which the rules of etiquette possess) and Mackie-Objectivity (which they lack) is troublesome.

Why, exactly, is something that is prescriptive but not Mackie-Objective naturalistically respectable? Mackie believes that facts about etiquette can be prescriptive and naturalistically respectable because their prescriptivity only requires the obtaining of a *causal connection* between the desired ends and the (best available) means for satisfying these ends.<sup>145</sup> What it is for something to be prescriptive is for it to consist in an appropriate causal relation between means and ends. And causal connections aren't queer.

An important qualification is that the 'ends' in question can be given not only by *individual desires of agents* but also by societies at large. Indeed, Mackie thinks that a stranger's sufferings "constitute some reason"<sup>146</sup> for us to help her so long as we think of this as a reason 'supplied by' the *societal* end not to let people suffer. So if we think that a stranger's suffering constitutes a reason to act then

we are again bringing in the requirements of ... an established way of thinking, a moral tradition, demands that I show some concern for the well-being of others<sup>147</sup>

---

<sup>144</sup> The examples I give, here and in other chapters, are examples of what Brink (1992: 9; also see Shafer-Landau 2003: 210; Nesbitt 1977: 223) calls 'mere etiquette', which is to be distinguished from those rules of etiquette that overlap with the rules of morality (examples of the latter are rules that oblige one to speak decently to other people).

<sup>145</sup> P28.

<sup>146</sup> P78; also see Joyce (2011a: 524); Svavarsdóttir (2001: 149).

<sup>147</sup> P78, 79.

Mackie even says that such *institution-dependent* moral language is “written into ordinary moral language.”<sup>148</sup> This however begets the question: doesn’t this compromise the error theory we started with? After all, we already had institution-independent moral language that is flawed because objectively prescriptive and written into ordinary moral language. But now we also have institution-*dependent* moral language that is not flawed because it is *subjectively* prescriptive and *also* written into ordinary moral language. The problem is that this opens up the possibility that the institution-dependent moral talk is *more important* than the institution-independent talk, in which case moral discourse is *not* in error (or at least its most important part is not in error). Mackie, however, doesn’t notice the threat.

Be that as it may, Mackie’s account faces another problem as well. Perhaps *motivational force* can be understood in terms of a causal connection between ends and means. But given that the reduction is supposed to be one of *Mackie-Prescriptivity*, which also involves normativity, a good question is whether a reduction of *normativity* to a causal relation between means and ends can work as well or whether we instead have an *elimination* of normativity. Offhand, it seems that we have an elimination. We have a strong intuition that normative relations outstrip whatever contingent causal relations can obtain between means and ends, even when sheep dog trials and etiquette are concerned. In the language of reductions introduced in §1.2.1, Mackie’s reduction is very reductionist in substance—so much so that it disallows us to retain one very important commonsensical thought about the property of normativity to be reduced; viz., that it can outstrip the contingently obtaining causal relations between means and ends. The worry is that this should make us seriously doubt the success of Mackie’s attempt to show that subjective prescriptivity is metaphysically kosher.

To sum up this sub-section, I have argued for three main claims. First, Mackie’s arguments from queerness, disagreement, and projectivism are unable to justify his Substantive Thesis. Second, if we can and must believe his Substantive Thesis then moral statements are all false, albeit not necessarily so. Third, subjectively prescriptive values exist, according to Mackie, although we have also seen reason to doubt this.

---

148 p79.

But again let's assume that Mackie's arguments work and that we should believe his error theory up until this point—hence that we should believe both its Non-Negotiable Commitment Claim and its Substantive Claim and that the theory can avoid the Formulation Dilemma. Let's ask, then, how Mackie recommends we should respond to the truth of moral error theory.

#### 2.2.4 Mackie's Solution

According to Mackie, what should we do with our error-ridden moral discourse once we realise that his error theory is true? Here the 'should' is a prudential, non-moral 'should' to avoid the charge of self-refutation.

Mackie's commentators come to very different conclusions about how Mackie answers this question. Some think that Mackie was a fictionalist. Fictionalists either think that we should *make-believe* rather than *believe* moral propositions (where to make-believe a proposition is to adopt an attitude akin to pretence to moral propositions so that one's attitude isn't subject to general norms of doxastic attitude adoption, such as the correspondence-truth norm). Or they think that we should continue to fully believe propositions as long as we have turned them into *fictionalist* propositions like: '*in the moral fiction*, stealing is wrong'.<sup>149</sup> Others think that Mackie was a preservatist, thinking that we should simply continue to moralize as if we had never discovered the error.<sup>150</sup>

I believe that all these interpretations are false and that Mackie was a *revolutionary cognitivist*. Mackie put forward a proposal for a *revolutionary reform* of our error-ridden moral concepts, and because he thought that moral concepts contribute to the meaning of moral statements he believed that moral statements and their meaning are to be changed as well.<sup>151</sup> So Mackie wasn't a preservatist because he thinks that we should *alter* rather than *preserve* moral discourse. And he wasn't a fictionalist because the kind of change that's needed neither involves pre-fixing

---

<sup>149</sup> Joyce (2005: 288); Sainsbury (2010: 204); Lillehammer (2004: 105). Further specification of moral fictionalism will be provided in Chapter 3.

<sup>150</sup> West (2010: 184-5); Oddie and Demetriou (2010: 200). This reading gains support from Mackie's remark that first-order and second-order ethical questions "are not merely distinct, but completely independent" (p16). For if they are then even if error theory is true we can still continue to engage in first-order moral discourse without falling into error.

<sup>151</sup> Schiffer (1990: 614) comes close to this interpretation.

moral propositions with a fictionalist operator nor taking some kind of make-belief stance to moral propositions. Instead, we should change moral concepts into *schmoral* concepts to get a *schmoral* statement like this:

**Schwrong**      Stealing is schmorally wrong

Participants in moral discourse should cease to assent to moral propositions that falsely suggest that type or token events exemplify objective prescriptivity and they should start to assent to *schmoral* propositions that *truly* suggest that type or token events exemplify *subjective* prescriptivity.

One passage that speaks in favour of my interpretation of Mackie's solution comes from a book that Mackie wrote a year before his *Ethics*:

[a] similar conceptual reform [to the one about personal identity I am presently proposing], rather than mere analysis of our present concepts, is, I believe, needed in ethics. I hope to discuss this topic in another book<sup>152</sup>

The 'conceptual reform' Mackie talks about is that of altering the contents of moral concepts from objective moral prescriptivity to subjective prescriptivity. The book he talks about is his *Ethics*.

Here is Mackie's idea presented in more detail. Distinguish between morality in the broad sense and morality in the narrow sense.<sup>153</sup> Morality in the broad sense is a general, all-inclusive theory of conduct which provides rules for behaviour reflecting objectively prescriptive moral truths. It also aims to realize moral ideals. Call the practice of discovering and communicating rules of morality in the broad sense *moralizing*. Morality in the narrow sense is limited to constraints on the extreme pursuit of self-interest. Its rules do not reflect objectively prescriptive moral truths and it doesn't aim to realize moral ideals. Morality in the narrow sense also trusts the agents' moral sense, is *man-made* ('invented' rather than 'discovered'—hence the subtitle of Mackie's *Ethics: inventing right and wrong*) and depends on

---

<sup>152</sup> Mackie (1976: 196n27); cf. Lillehammer (2004: 106); Silverstein (2012: 13). Burgess is therefore off the mark when he writes that: "even though [Mackie] talks of the need to invent morality, he does not seem to think that this proposal could be worked into a revisionary meta-ethic" (1998: 538n8; also see Blackburn 1993: 149). He thinks it can, and proceeds to show this in the later chapters of his *Ethics*.

<sup>153</sup> Mackie (1977: 106-16).

individual and collective benefits, which are subject to change over time.<sup>154</sup> Call the practice of finding and communicating moral rules in the narrow sense *schmoralizing*. Mackie's proposal is that we should stop moralizing and start schmoralizing.

How can schmorality be implemented—that is, how can we make sure that people are no longer in error when they think and talk about morality and start to think and talk about schmorality, and thus avoid the error? In the chapters that come after his rejection of moral values Mackie seemingly simply continues to talk about moral values. This, I suspect, has fuelled preservatist readings of his *Ethics*. I think that what Mackie had in mind though is that we cease to believe that there are moral properties when we say 'stealing is wrong' and instead start to believe that there are *schmoral* properties when we say 'stealing is wrong'.<sup>155</sup> This entails that it would be more accurate to say 'stealing is *schmorally* wrong', but perhaps because Mackie thinks that the folk couldn't be brought to start *talking* differently he allows them to continue to *say* 'stealing is wrong'. This is fine, of course, as long as we know that what we mean when we say that has changed.

What is the function of *schmorality* and of believing in *schmoral* propositions? It is something like securing the benefits of living in a well-run society and enjoying the fruits of cooperation.<sup>156</sup> We can make this more precise by looking at what Copp calls the *problem of sociality*:

**Problem of Sociality**

Humans are typically not self-sufficient and can benefit from social cooperation. But the problem is how we can guarantee the benefits of social cooperation given that it is often in humans' immediate, short-term self-interest not to cooperate<sup>157</sup>

---

<sup>154</sup> Lillehammer (2004: 96) argues that Mackie's morality in the narrow sense can be interpreted as a form of morality that already exists and that can be shown to be in error. This is a mistake. The folk wouldn't mistake 'their' Mackie-Objective and Mackie-Prescriptive morality for morality in the narrow sense—the latter is 'man-made' and therefore not "prior to and logically independent of all [human] activities" (Mackie 1977: 30).

<sup>155</sup> P106.

<sup>156</sup> Also see, e.g., Gauthier (1986).

<sup>157</sup> Copp (2009a: 27).

For Mackie, the function of schmorality is to counteract these limited sympathies that stand in the way of solving the Problem of Sociality.<sup>158</sup>

This begets the question how Mackie can explain how schmoral constraints on selfish behaviour ‘bind’ agents in the absence of Mackie-Objective and Mackie-Prescriptive moral values.<sup>159</sup> After all, if we can simply shrug off schmoral constraints whenever we are presented with them because they are only *subjectively* prescriptive (alter your ends, and you will have altered what you ought to do—recall that subjective prescriptivity consists in causal relations between means and *ends*) then they aren’t really *constraints*. For consider the following conversation

**Abe:** ‘Hey, don’t steal those jeans! That is “schmorally” wrong!’

**Bert:** ‘So what? There isn’t any desire-independent normativity that speaks against my stealing these jeans and there also isn’t anything in the world that affects my motivation. Moreover, I personally desire to steal, this act of stealing won’t disrupt society and no-one’s looking (so the risk of getting caught and ending up in jail is low). So why wouldn’t I?’<sup>160</sup>

As we defined schmoral constraints, this is a perfectly legitimate way for Bert to respond. But then it no longer looks as though the constraints are really *constraints*. And so it seems that we can’t solve the Problem of Sociality with schmorality. Implementing a ‘schmorality’ with the aim of solving the Problem of Sociality is likely to be unsuccessful.

Mackie thinks he has a reply to this. It isn’t that calculations of long-term self-interest as brought out by various game-theoretic considerations can do the trick.<sup>161</sup> Instead he argues that what we need are “psychological substitutes for physical chains.”<sup>162</sup> Mackie’s proposal has two components. First we should “trim down moral demands to fit present human capacity” and “look for rules ... that can fit with the relatively permanent tendencies of human motives and thought.”<sup>163</sup> This

---

<sup>158</sup> P107.

<sup>159</sup> Blackburn (1985).

<sup>160</sup> Adapted from Hare (1999: 4).

<sup>161</sup> P114, 119.

<sup>162</sup> P116.

<sup>163</sup> P133, p134. The talk of ‘physical chains’ further corroborates my interpretation of Mackie’s Non-Negotiable Commitment Claim, according to which for Mackie

helps to solve the problem of getting actual constraints without objective prescriptivity. For if there are fewer cases where what one schmorally ought to do is out of line with what one wants to do, then, on the assumption that a successful reduction of non-objective prescriptivity to the obtaining of a causal relation between means and ends works, there will be fewer cases where we have to say that the schmoral constraint in fact wasn't doing any constraining. But we still need something in addition to this for those cases where people want to do things (such as stealing) that inhibit social stability and cooperation. (Moreover, as I argued in the previous section, it is not at all clear that the reduction of non-objective prescriptivity to causal means-end relations works).

Here is where the second component of the proposal comes in. If you don't desire to do what is schmorally right (i.e., foster social cooperation) then there is a real *penalty* that awaits you. The penalty is a disintegrated psychology and the sacrificing of "inward peace of mind".<sup>164</sup> The main problem with this proposal, of course, is that for many people a 'disintegrated' psychology, it seems safe to fear, wouldn't be enough of a penalty to forego stealing (or worse) when they are sure that no one is looking. But if that is true then schmorality won't solve the problem of sociality. Schmorality's 'constraints' on purely selfish behaviour aren't really *constraints*. Perhaps, then, it is better after all to be a fictionalist or a preservatist—but these are options that Mackie doesn't consider.

To sum up, I have defended the claim that the best interpretation of Mackie's proposal for what to do with moral discourse after error theory is that it is a *revolutionary cognitivist* proposal. According to revolutionary cognitivism, ordinary participants in moral discourse should cease believing moral propositions and start believing *schmoral* propositions, where *schmoral* propositions are propositions like 'stealing is schmorally wrong' and whose truth does not entail objective prescriptivity but only the defensible kind of subjective prescriptivity. I have also expressed reservations about whether Mackie's system of *schmoral* obligations will work for the purpose for which it was invented; viz., solving the Problem of Sociality with the particular aim of safeguarding social cooperation.

---

the fate of morality hinges on moral properties with actual 'binding' or 'motivating' power.

<sup>164</sup> P191. The quote is from Hume (1751: 82.)



This means that Mackie doesn't have a working theory about moral discourse that we can use after error theory. Moreover, Mackie doesn't have convincing arguments for this Non-Negotiable Commitment Claim or his Substantive Claim. Finally, Mackie's error theory doesn't have the resources to respond to the Formulation Dilemma. Overall, then, Mackie solves none of the problems for moral error theory as discussed in §1.4. Let's see if the other main well-worked error theory currently discussed in the literature—that of Richard Joyce—does better than this.

## 2.3 Richard Joyce

### 2.3.1 Joyce's Non-Negotiable Commitment Claim

According to Lillehammer, Joyce's error theory contrasts with Mackie's in that the former, but not the latter, is based on a Kantian style focus on categorical reasons.<sup>165</sup> I argued above that this isn't entirely accurate insofar as Lillehammer's interpretation of Mackie is concerned. Mackie believed that talking about categorical reasons, furnished by categorical *imperatives*, make his case against moral values "clearer."<sup>166</sup> What is true is that Mackie's main focus is on moral values whereas Joyce focuses squarely on categorical reasons as the non-negotiable commitment of moral discourse.

What, exactly, do we mean by *categorical*? Consider the following from Dreier:

First, we might say that an imperative is categorical when our application of it to the behaviour (or deliberations) of someone does not depend on any aim, or any desire of that person. In this sense, the rule 'Practice your scales daily' is not categorical, since we should withdraw it upon learning that the addressee had no interest in learning to play the piano. There is little question that moral imperatives aren't categorical in this sense. Informing your critic that you aren't interested in according respect to other persons isn't going to make him withdraw the imperative to keep your promises. [There is also] a second sense of categoricity. A rough try at

---

<sup>165</sup> Lillehammer (2004: 101).

<sup>166</sup> Miller (2013: 108) uses the locution "furnished by." The Mackie quote is from P27.

expressing this sense is to say that each person has reason to follow, no matter what her desires are<sup>167</sup>

For Joyce, morality and especially its obligations are categorical in both senses. Moral obligations *apply* to all of us regardless of our desires, wants or goals. You can't escape being under a moral obligation by changing your desires in the way that you can escape being under an obligation to practice your scales by changing your desire to learn how to play the piano. Say when this is true that morality is *categorically applicable*. Moral obligations also supply all of us with *reasons* to act, again regardless of our desires, wants or goals. You can't escape having a reason to abide by your moral obligation by changing your desires whereas you can escape having a reason to practice your scales by changing your desire to learn how to play the piano well. Say when this is true that morality is *categorically reason-giving*. Importantly, these 'reasons' are standard normative reasons or reasons of rationality—they are not 'merely' reasons with moral content which it would not be irrational to ignore. For ease of exposition, and when context allows me to, I will sometimes treat this last qualification as read from now on.

In this terminology, morality's categorical reason-giving force entails its categorical applicability: if everyone has a reason to abide by morality's rules then, straightforwardly, the rules also apply to everyone. But some philosophers argue that the reverse entailment does not hold.<sup>168</sup> Joyce disagrees. For Joyce thinks that *Mackie's platitude* is true:

---

<sup>167</sup> Dreier (1997: 84); also see Railton (2010: 268); Brink (1997). Railton (1986: 165) also defines categoricity in terms of all agents necessarily having reason to obey moral imperatives (he reserves the term 'universality to talk about what Dreier defines as the first sense of 'categorical'). To see the gap between both sense of categoricity, consider, with Foot (1972: 308), the case of etiquette, where we use the 'should' of etiquette categorically but "to which no one attributes the special dignity and necessity conveyed by the description "categorical imperative."" For a rule of etiquette "does not fail to apply to someone who has his own good reasons for ignoring this piece of nonsense, or who simply does not care about what, from the point of view of etiquette, he should do." Also see Shafer-Landau (2003: 199).

<sup>168</sup> Foot (1972); Kavka (1984); Brink (1989: Chapter 3, 1992); Railton (1986); Copp (1995). Foot "later officially retracted her heresy" (Finlay 2008a: 349).

### Mackie's Platitude

[I]t is necessary and *a priori* that, for any agent  $x$ , if  $x$  ought to  $\phi$ , then  $x$  has a reason to  $\phi$ <sup>169</sup>

From Mackie's Platitude it follows that we can move from ' $x$  morally ought to  $\phi$ ' to ' $x$  has a reason to  $\phi$ ' on the ground that it is a *platitude* that oughts entail reasons. That is, just like it is a platitude that being red entails being coloured, it is a platitude that oughts entail reasons.

Shafer-Landau complains that Mackie's platitude isn't a platitude.<sup>170</sup> No one contests the platitude that red objects are coloured whereas lots of philosophers (and perhaps other folk) doubt that morality is normative in the strong sense of entailing categorical reasons. Shafer-Landau argues that the contestability of Mackie's 'platitude' is evidence that it isn't in fact a platitude, whereas the non-contestability of the platitude that all red objects are coloured shows that this platitude in fact is a platitude.<sup>171</sup> It may still be true, of course, that 'for any agent  $x$ , if  $x$  ought to  $\phi$ , then  $x$  has a reason to  $\phi$ ', but, argues Shafer-Landau, that won't be because it is *platitudinously* true. We need a different argument for why oughts entail reasons.

Shafer-Landau is right. Platitudes need not be recognized as platitudes by *absolutely everyone* to *count* as platitudes. I might be too tired to recognize that something is a platitude. But platitudes do need to be recognized by enough people in the debate to count as platitudes. Especially people like Brink, Railton and Copp who have thought long and hard about the issue of morality's normativity shouldn't (be able) to deny the alleged platitude that oughts entail reasons. The mere fact that *they* deny this is evidence that Mackie's Platitude isn't a platitude. In light of this it seems question-begging to insist that Mackie's platitude is a platitude, so Joyce should have further arguments for morality's commitment to categorical reasons. He does, and I will now explain what these are.

Joyce accepts that it would have to be part of a moral concept that there are categorical reasons—Joyce accepts what I have called the Conceptual Entailment

---

<sup>169</sup> Joyce (2001: 101). Also see Broome (2007).

<sup>170</sup> Shafer-Landau (2005: 101).

<sup>171</sup> Shafer-Landau (2005: 111).

Claim. His commentators widely agree: “Joyce claims it is a conceptual truth that moral obligations have “inescapable authority.”<sup>172</sup> The reference to ‘inescapable authority’ (sometimes Joyce also uses ‘practical clout’ or ‘practical oomph’) is important. Joyce thinks that what he can establish is that when the folk are engaged in moral discourse they are committed to something we can only rather inchoately formulate as the ‘inescapable authority’ of moral considerations. The folk make inchoately formulated claims like you ‘have to’ refrain from stealing and killing ‘no matter what you want’.<sup>173</sup> This is the ‘bare’ conceptual entailment claim: it is part of the correct application conditions on moral concepts that you apply them to things that there is inescapable authority to do or refrain from doing. Furthermore, Joyce thinks that our best *philosophical attempts at interpreting* this notion of inescapable authority commits us to talking about categorical *reasons* of rationality.<sup>174</sup>

Why do we end up with categorical reasons of rationality when we try to get clearer on the notion of inescapable authority? The reason is that categorical reasons of rationality are authoritative and inescapable in the very same way that moral obligations are meant to be authoritative and inescapable. Morality delivers obligations (morality is authoritative) in such a way that there is nothing the agent can do, not even adopting different desires, that allows him to escape this obligation (morality is inescapable). Similarly, practical reasons too are authoritative (they ‘tell you’ what to do) and inescapable. For whenever you try to question their authority you are *ipso facto* committed to their authority because to question something is to

---

<sup>172</sup> Copp (2010: 155); also see Joyce (MS). In fact moral discourse exhibits two conceptual commitments, according to Joyce: “The conceptual commitment of moral discourse, upon which the error theory runs, is one concerning actions that we “have to” perform, regardless of what our desires and interests are. We can place next to this another commitment concerning the content of such imperatives: that the things we “have to” refrain from need to include breaking promises for trivial reasons, doing violence for self-gain, and the like. ... In other words, a theory of imperatives that managed to supply strong categorical imperatives ... but for things like “Kill anyone who annoys you,” “Steal when you can,” etc., simply would not be a morality” (2001: 67). Joyce, as noted, is happy to leave matters imprecise regarding the second entailment. I follow suit because the main beef in the argument is that it is part of moral concepts’ correct application conditions that they apply to things that exhibit inescapable authority.

<sup>173</sup> Joyce (2001: 76). Also see Joyce (2000).

<sup>174</sup> Joyce (2001: 104; 2006: 192). Although Joyce also has his doubts (2009a, 2000b).

ask for a reason.<sup>175</sup> This only follows, as Joyce realizes, when reasons are given by practical rationality. He writes: “with Smith [1994] I will use “normative reason” in a restricted sense, to mean something that is justified according to practical rationality.”<sup>176</sup> Joyce also believes that whether agents have a reason to do something is determined by whether they would want to do that thing if they reasoned correctly—that is, if they reasoned without flouting such principles as that which obliges agents to take the means to their ends.

Moreover, it looks as though practical rationality is *uniquely* inescapable. What else has this feature of being authoritative *and* inescapable? Nothing, it seems. Therefore, the only hope for defenders of moral realism would be to show that there are categorical practical reasons with recognizable moral subject-matter. (To foreshadow what is to come: Joyce’s Substantive Claim is that there are no such reasons.)

Let’s look at this argument, and the considerations Joyce brings to the table in support of its premises, in more detail. Joyce’s master-argument looks like this:

#### **Joyce’s Master Argument**

- P1     If x morally ought to  $\phi$ , then x ought to  $\phi$  regardless of whether he cares to  $\phi$ , regardless of whether  $\phi$ ing satisfies any of his desires or furthers his interests—morality has categorical applicability
- P2     If x morally ought to  $\phi$ , then x has a reason for  $\phi$ ing—morality has categorical reason-giving force
- C1     Therefore, if x morally ought to  $\phi$ , then x has a reason for  $\phi$ ing regardless of whether  $\phi$ ing serves his desires or furthers his interests [From P1, P2]
- P3     But there is no sense to be made of such reasons
- C2     Therefore, x is never under a moral obligation [From C1, P3]<sup>177</sup>

In this sub-section I discuss Joyce’s arguments for (P1-P2). The next sub-section discusses Joyce’s argument for (P3). The sub-section after that discusses Joyce’s moral fictionalism, which is Joyce’s view on what we (prudentially) ought to do with moral discourse after error theory.

---

<sup>175</sup> Joyce (2001: 49, 68, 102, 104). Also see Shafer-Landau (2005: 110). Copp (2007: 271) calls this the closed question proposal of authoritative normativity (or inescapable authority).

<sup>176</sup> Joyce (2001: 70 n. 21).

<sup>177</sup> Joyce (2001: 42).

Let's start with three dialectical points about P1. First, P1 can be—and usually is—accepted by lots of naturalist theories that Joyce has no beef with other than that they fail to recognize morality's *categorical reason-giving force* as mentioned in P2.<sup>178</sup> For as we have seen, there is only an entailment from categorical reasons to the categorical applicability of moral norms; there isn't one that runs in the other direction. So these naturalists accept P1 but reject P2. They either think that an honestly executed internal accommodation project doesn't commit us to P2 (non-revisionary naturalism). Or else they think that when it does that a revision of our conception of what morality would have to look like to one that doesn't include categorical reasons would be *benign* and thus vindicates moral discourses (revisionary naturalism).

Second, P1 is very plausible. Joyce brings this out with various examples. For instance, we have the intuition that the rules of morality apply to characters like Plato's Gyges, *full stop*. Gyges has a ring that makes him invisible at will, and he uses it to steal and kill as he can always get away with it by making himself invisible. We think that it doesn't matter what Gyges desires to do, or whether he can escape punishment, or what have you. Gyges shouldn't steal and kill, *period*.<sup>179</sup> Again, the thought that *morality* forbids Gyges to use his ring in the ways he wants is very much ingrained in our ordinary moral thought (and Joyce exploits this well); the beef of Joyce's argument is on the issue of whether the 'ought' in 'Gyges 'ought' not to steal and kill' must be analysed as a categorical normative reason.<sup>180</sup>

Nevertheless, third, P1 *has* been denied. Consider Harman's moral relativism, according to which the 'logic' of at least the deontic moral judgments forces us to conclude that they can only sensibly be applied to those agents whom we know are capable of being motivated by the reasons we ascribe to them.<sup>181</sup> If this is true then P1 is false: 'if x morally ought to  $\phi$ , then *it is not the case that* x ought to  $\phi$  regardless of whether he cares to  $\phi$ , regardless of whether  $\phi$ ing satisfies any of his desires or furthers his interests.' However, Harman's view is very much a minority view and

---

<sup>178</sup> These naturalists include: Brink (1984, 1989); Railton (1986); Foot (1972).

<sup>179</sup> Joyce (2001: 32-3, 2011a: 525). See Plato, *Republic*, 360b-c.

<sup>180</sup> Also see Joyce (2006: 203).

<sup>181</sup> Harman (1975).

[t]he many critical howls that greeted [Harman's] claim are evidence that [P1] was regarded by his audience as a conceptual truth<sup>182</sup>

Moreover, I will have my own arguments against Harman in Chapter 5. So here I proceed on the assumption that Joyce has settled P1. We thus accept that “moral commands are conditioned on the desires and concerns of those to whom they are addressed neither explicitly nor implicitly.”<sup>183</sup>

It is much harder for Joyce to argue that P2 is true. Apart from the failed entailment between P1 and P2 based on Mackie's platitude, Joyce presents the *translation test argument* and reinforces that argument with the *function* or *use argument*. I discuss both of them in what follows.

According to the translation test argument, if you have a concept x (such as MORAL WRONGNESS) and wonder whether it entails concept y (such as INESCAPABLE AUTHORITY), then you should remove y from your discourse about x and see if the new discourse would be accepted as an eligible replacement of the original discourse.<sup>184</sup> If the answer is ‘no’ then x entails y.

A big problem with the translation test argument is that it isn't feasible. For the counterfactual judgments that are involved in it are too complex for us to trust our intuitions about them.<sup>185</sup> Can we get a clear verdict on whether a certain contender moral concept counts as a moral concept in the hypothetical situation that we are confronted with a choice between it and some other contender moral concept? Problems of projecting theoretical commitments onto moral discourse loom large, as does the problem that the folk may lack the conceptual sophistication to grasp the question at hand. So even though the argument might work in theory, in practice it is pretty much useless. Or at least this is what Joyce himself believes, and because I want to argue that Joyce's arguments for error theory are highly problematic I'll follow him in this.

---

<sup>182</sup> Shafer-Landau (2005: 118n4).

<sup>183</sup> Joyce (2011a: 523).

<sup>184</sup> Joyce (2001: 3, 2006: 200); also see Copp (2007: Ch. 6); Tresan (2010: 227).

<sup>185</sup> Joyce (2001: 3, 2006: 201).

So the translation test needs reinforcement. This is what the *use argument* brings.<sup>186</sup> If we ask what determines whether something like INESCAPABLE AUTHORITY is part of moral concepts then according to Joyce

[T]he answer turns on how the concept *morality* is *used*. If concept A is used in a certain manner, but turns out to be problematic for various reasons (i.e., it is uninstantiated by the world), and concept B is an instantiated contender for replacing A, then B can be an adequate successor only if it too can be used in the same manner. For example ... when we discovered that there are no diabolical supernatural forces in the universe, we had no further use for the concept *witch*. Perhaps we could have carried on applying the word “witch” to women who play a certain kind of local cultural role on the margins of formal society ... but carrying on in this way would not have allowed us to *use* the word “witch” for the purposes to which we had previously put it: to condemn these women for their evil magical influence and justify their being killed<sup>187</sup>

Likewise for moral concepts, argues Joyce. If taking inescapable authority out of (the correct application conditions of) moral concepts then we won’t be able to use moral concepts in the same way that we can use them now; therefore, inescapable authority is part of (the correct application conditions of) moral concepts. So on the basis of the use argument we can formulate:

**Practical Debunking Thesis (PDT)**

The usefulness of moral statements does not survive the debunking of the myth that moral concepts have inescapable authority; therefore, inescapable authority is an essential part of moral concepts<sup>188</sup>

I argue that Joyce uses two types of use arguments: one that focusses on the function of moral concepts as used by agents in their own practical deliberation, and another that focusses on the ability of agents to criticize the behaviour of other agents whilst using those concepts. I call the former the *silencing function* argument.<sup>189</sup> I call the latter the argument *from enfeebled moral criticism*.<sup>190</sup> The

---

<sup>186</sup> Joyce (2007b, 2006: 206-8); also see Mackie (1977: 108, 203), Cuneo (2012: 114, 124); Tresan (2010: 235-6); Garner (1994: 139n7).

<sup>187</sup> Joyce (2007b: 65).

<sup>188</sup> Tresan (2010: 227).

<sup>189</sup> Joyce (2001: 181, 215; 2006: 108-18, 168-74, 206); Kirchin (2010: 179); Cuneo (2012: 115).

<sup>190</sup> Joyce (2006: 208); Tresan (2010: 232); Vogelstein (forthcoming).



literature further distinguishes the *redundancy of moral language* argument and the argument *from encouraging wrongdoing*.<sup>191</sup> I explain after my own interpretation of Joyce's use arguments why both of these arguments merit only a very brief discussion.

Start with the silencing function argument. According to this argument, it is a *sine qua non* for moral concepts that (i) when they feature in moral judgments like 'stealing is wrong' or 'I ought not to steal', they *have the following effect on the agent* who makes those judgments. They put out of mind rationalizations that might tempt them to steal. According to the silencing function argument, it is also true that (ii) moral concepts could only have this effect when the moral facts they can be used to describe are facts about categorical moral reasons. In other words, (i) the *function* of moral concepts is to put alternative rationalizations for the performance of different actions out of the agent's mind and (ii) moral concepts could only perform this function if the facts they describe are facts about categorical reasons. Aspiring moral concepts that cannot be used to describe facts about categorical reasons are *schmoral* concepts at best.

One question that comes to mind is whether the silencing function argument entails that the normativity of categorical reasons must be *overriding* in the sense that once it is recognized it should and does, precisely, put other rationalizations—such as the excitement that stealing might bring or the gains to one's personal welfare—out of mind. That is, if it is allowed that non-moral (say, prudential) considerations can override moral ones, would it then follow that its countervailing consideration wasn't in fact a moral consideration? After all, the 'moral' considerations didn't perform its alleged function, which was to 'silence' these other considerations. If Joyce's argument assumes the overridingness of moral considerations' normative force then that is a strike against the argument. For it does not seem to be the case that the normativity of moral considerations is overriding in this way.<sup>192</sup> At least we need an argument that morality's normativity *has* this further property of being overriding. But this has not been supplied by Joyce.<sup>193</sup>

---

<sup>191</sup> Tresan (2010: 234-6) based on Joyce (2006: 206-7).

<sup>192</sup> For such arguments see, e.g., Kagan (1989: 1).

<sup>193</sup> Joyce comes close to accepting that morality has to be overridingly normative in his discussion of what we ought to do with moral discourse after recognition of the

To the contrary; Joyce seems to think that the overridingness of moral considerations plays no role in his silencing argument:

a moral judgment ... is no *guarantor* that the action will be performed, but so long as it increases the likelihood of the performance then this may be its evolutionary function [because the moral judgment can] ... play a dynamic role in deliberation .... prompting and strengthening certain desires and blocking certain considerations from even arising<sup>194</sup>

But this passage is unclear. Either moral considerations play a ‘dynamic role’ in deliberation in that although they ‘increase’ the likelihood of the agent performing the morally right action, they need not necessarily do that and hence may be overruled by other considerations. Call this the ‘dynamic role’-interpretation. Or else moral considerations play an even more impressive role by actually ‘blocking’ other considerations from even arising. Call this the ‘guaranteeing role’-interpretation. What is unclear about this passage is that it supports both interpretations. Let’s take each in turn.

The guaranteeing role interpretation of morality’s function is tantamount to the interpretation we had in the previous paragraph. Moral considerations only count as *moral* considerations if they override other considerations. But we have already seen problems with this suggestion. So let’s take the ‘dynamic role’ interpretation instead. The problem with this interpretation is that the weakened account of moral considerations’ functional role in an agent’s psychology that it implies makes it a lot harder to set them apart from other considerations. That is, if sometimes prudential considerations ‘silence’ rationalizations for doing a ‘morally right’ thing then we get the following question. How often do moral considerations have to be successful in silencing other considerations to count as *moral* considerations?<sup>195</sup> Without an

---

truth of error theory, although I’m not sure that this issue has completely crystallized in his thinking (2001: 184).

<sup>194</sup> Joyce (2006: 114).

<sup>195</sup> See Lewis (1989) for a similar question about how many commonly recognized aspects of the concept ‘value’ would have to be instantiated by the world for us to be able to say that there are values (perhaps benignly reduced to less-than-perfect claimants of the term of which it is true that the world satisfies all of its aspects but of which it is also true that this claimant doesn’t include all the commonly recognized aspects of values as we actually think about them). See Joyce (2012) for critique.

answer to that question it looks as though their ability to “silence” other considerations is insufficient to set them apart from other considerations, for these also have this silencing function. Note that responding by insisting that we can pick out moral considerations by looking at their content *as well* as their silencing function might well work, but also *obfuscates* the need for silencing. For if both moral and prudential considerations can silence each other and are therefore on a par with each other in that respect then constraints on the *content* of moral considerations are doing all the philosophical work.

Even if we assume that Joyce can solve this problem, there is a further problem with his silencing function argument. Certain naturalists can also claim that moral thought and discourse plays much this function without postulating categorical reasons. These naturalists can accept (i) but reject (ii). Cuneo gives the example of *pretending* that there are categorical moral reasons, which, when we are sufficiently *engrossed* in the moral game, will also tend to set us on the right path.<sup>196</sup> If we pretend that there are categorical reasons when we are doing our practical reasoning then, so long as we are sufficiently ‘down’ with the moral game, we will also manage to put out of mind rationalizations for different actions, and indeed very reliably so, without a need to assume that the moral concepts that feature in those considerations must describe facts about categorical moral reasons. So, the objection goes, even if moral considerations play a silencing function and even if we can somehow unproblematically describe what this function amounts to (recall my earlier comments about the overridingness of prudential considerations that suggest that this assumption isn’t at all obvious), the data as we have them do not entail the claim that moral considerations require categorical reasons.

The next point of critique I have about the silencing function argument is related to the previous one. I just argued that it is not clear why ethical naturalism wouldn’t be able to account for the datum that moral considerations play a silencing function. But we can also give Joyce the benefit of the doubt and say that ethical naturalism cannot give a good enough explanation of this datum. In that case still it is not clear that (ii) follows from (i); viz., that it follows from (i) that moral *concepts* would have to describe facts that are categorically reason-giving. Perhaps, that is,

---

<sup>196</sup> Cuneo (2012: 123-4). He bases himself on Johnston (2010: 17). Also see Brink (1989: 66).

moral statements stand in a more general relation of entailment to categorical moral reasons, and this relation is ‘more general’ because it doesn’t require connections between *concepts*. I call this relation ‘metaphysical entailment’ and explain it more fully in the next chapter. Because Joyce doesn’t consider other possibilities besides conceptual entailment he jumps from (i) to (ii) without argument. But in fact argumentation is required. So the dialectical value of this particular objection to the silencing function argument is that it may well be a good argument for error theory (as long as the other problems can be solved with it anyway) but that it is unclear whether it would be a good argument for an error theory that uses *conceptual* entailment as opposed to a good argument for error theory that uses *metaphysical* entailment (or some other account of the mode of commitment of moral discourse to the moral reality thesis for that matter). On the basis of this and the other problems discussed I conclude that the silencing function argument doesn’t make the Conceptual Entailment Claim sufficiently plausible.

Let’s move on to the next version of the use argument and see whether it can bring solace. This is the argument from enfeebled moral criticism. The argument starts with the observation that it would be odd to say of someone that she did the morally wrong thing but had no reason not to do it. The reason why this is odd is that we cannot legitimately criticize someone for doing something she had no reason not to do. But then holding constant the practice of moral criticism, moral considerations must entail reasons; and because moral criticism applies *categorically* we even get *categorical* reasons.

Vogelstein argues that the kind of criticizability at least sometimes has to be of the *subjective* and cannot always be of the *objective* kind.<sup>197</sup> It is extremely implausible that I’m culpable for the consequences of my behaviour that I’m unaware of at least where I’m not culpable for the unawareness (objective wrongness).<sup>198</sup> It is much more plausible that I’m culpable for the consequences of

---

<sup>197</sup> Vogelstein (forthcoming). He also argues that the concept of criticism here has to refer to the attitude of being critical rather than to the act of criticism, as actually criticising wrongdoers might make things worse. Exactly what kind of worseness is at stake here however remains unclear, and if Vogelstein’s suggestion is that criticising a wrongdoer will make things *morally* worse than that suggestion will be immediately rejected by the error theorist. So I’ll leave this out.

<sup>198</sup> This further condition is not recognized by Vogelstein but it is clearly required for his point to go through.

my behaviour that I am aware of (subjective wrongness). Taking this into account, the argument from enfeebled moral criticism will go something like this:

**Argument from Enfeebled Moral Criticism**

- P4     Moral practice is committed to holding transgressors blameworthy regardless of the ends they desire
- P5     Holding someone blameworthy regardless of the ends they desire entails judging that that person failed to respond to reasons that were authoritative for her<sup>199</sup>
- C3     Therefore, moral practice is committed to the judgment that moral reasons are authoritative for persons whatever their ends [From P4, P5]<sup>200</sup>

What to make of this argument? P4 seems true. Barring, perhaps, genuine moral dilemmas such that both  $\phi$ -ing and not  $\phi$ -ing would be morally wrong, it seems that doing something morally wrong deserves (moral) criticism. And since moral obligations apply to agents categorically, the critique of their behaviour that we should be able to direct at them also takes no regard of these agents' desires or ends.

But it is not clear that P5 is true. Generally speaking, if you are criticisable for not  $\phi$ -ing (refraining from acting is also an action) then there must have been *some consideration that spoke against*  $\phi$ -ing. The thought is this: it would be very weird if we could criticise you for doing something if there wasn't *anything at all* that counted against you doing that thing. But whence the focus on *reasons of rationality* that do the counting against? Why wouldn't a weaker conception of the normativity of morality (as opposed to the strong, reason involving-conception) be enough? One example of a weak conception of morality's normativity is a *rule-conception* of normativity:

---

<sup>199</sup> Also see Nagel (1970: 83); Shafer-Landau (2003: 193).

<sup>200</sup> Taken, with adaptations, from Finlay (2008a: 358). Vogelstein (forthcoming; minor adaptations mine) has: (P4') Necessarily, for all A, x: if it would be morally wrong for A to do x, then A would be criticisable if she were to do x; (P5') Necessarily, for all A, x: if A would be criticisable if she were to do x, then if A were to do x, A would not be in compliance with her best (subjective or objective) reasons; (C3') Necessarily, for all A, x: if it would be morally wrong for A to do x, then if A were to do x, A would not be in compliance with her best (subjective or objective) reasons [From P4', P5'].

### **Rule-Involving Conception of Morality's Normativity**

Moral oughts are normative in virtue of being rules, and rules are normative in virtue of distinguishing between what is *correct* and *incorrect* or what is *allowed* and *disallowed*<sup>201</sup>

If morality can be normative simply by consisting of rules and thus without (necessarily) involving reasons then there can be something that counts against not  $\phi$ -ing—viz., morality's rule that tells you not to  $\phi$ —in the absence of rationally authoritative reasons. The rules of morality apply categorically, and there is genuine normativity around (albeit it of the rule- rather than rationality-involving kind), but P5 will be false. Holding someone blameworthy regardless of the ends they desire does *not* entail judging that that person failed to respond to reasons that were authoritative for her. It may only entail that the agent ignored categorically applicable normativity that falls short of being rationally authoritative; for instance, a kind of normativity that can be analysed in terms of rules.

Joyce disagrees with this last point, and tries to elicit precisely that intuition in us with cases like:

#### **Jack the Ripper**

It would be intolerably odd if it were possible to add to our observations of Jack the Ripper's atrocities that he "had every reason to act wickedly ... and no real reason to refrain",<sup>202</sup>

I accept that it would be intolerably odd if there were *nothing* we could say to Jack the Ripper. But with the rule-involving rather than reason-involving conception of morality's normativity we can say *something*: we can say that what he did was wrong *relative to the rules of morality* and that since rules are categorically normative, Jack's behaviour is 'categorically wrong' and we can criticize him for ignoring this rule. Now perhaps this would still be enfeebled moral criticism—if

---

<sup>201</sup> Parfit (2012 Vol1: 144). One difference between the rule-involving conception of normativity and the reason-involving conception of normativity is that on the former conception, new normative truths can be created "merely by introducing, or getting some people to accept, some rule. Legislators can create laws, and anyone can create the rules that define some new game ... In contrast, on the reason-involving conception, there is normativity only when there are normative reasons ... We cannot create such reasons merely by getting people to accept some rule" (*ibid.*).

<sup>202</sup> Based on Joyce (2006: 204); also see Joyce (2001: 43-4). Jack the Ripper was a serial killer in London in the late 1880s.

inescapable authority isn't an essential part of moral concepts then we can't criticise Jack for ignoring the normative *reason* (he knew) he had not to kill and we can only point to a rule—but the question is whether this degree of enfeebling moral criticism is *sufficient* for the practical debunking thesis.<sup>203</sup> If the choice is between moral concepts with inescapable authority that allow us to criticise agents for ignoring reasons (they knew they had) to be moral and moral concepts that don't allow us to criticize immoral behaviour *at all*, then we might be drawn to the practical debunking thesis. But if the choice is between moral concepts with inescapable authority that allow us to criticise agents for ignoring reasons (they knew they had) to be moral and moral concepts that allow us to criticise agents for wrong behaviour relative to the rules of morality, then it is not at all clear that *this* enfeebling of our capacity to criticize wrongdoers vindicates the practical debunking thesis. The *practical vindication thesis* (PVT) might still be on the table: the usefulness of moral statements survives the debunking of the myth that moral concepts have inescapable authority, so inescapable authority is not an essential part of moral concepts. Joyce doesn't consider this issue, and so, again, his arguments are not conclusive.

A further point of critique I want to make regarding the argument from enfeebled moral criticism is structurally identical to a remark I made a couple of paragraphs up concerning the silencing function argument. The remark I have in mind was that the silencing function argument might be a good argument for *a kind of error theory* but that it is not clear that it is a good argument for error theory *based on conceptual entailment*. For a more general entailment relation was still in the running, and nothing about the silencing function argument seemed to rule out that entailment relation. The same holds for the argument from enfeebled moral criticism. Although I just gave reasons to doubt this, let's assume that Joyce is right that moral discourse wouldn't be *moral* discourse without the possibility of moral criticism and that we need categorical reasons of rationality for moral criticism. But why would it follow from this that it is built into the (correct application conditions of) moral *concepts* that the things they can be used to describe are categorical reasons? Again, a general entailment relation, which I call 'metaphysical entailment' and which will be further elucidated in the next chapter, would seem to do just as well. Joyce doesn't consider this option; probably, as I surmised earlier as well,

---

<sup>203</sup> Again cf. Lewis (1989); Joyce (2012).

because Joyce is working with only one conception of entailment: conceptual entailment. My critique is that Joyce hasn't done enough to argue that the argument from enfeebled moral criticism supports the *conceptual* entailment. There are too many problems with this version of the use argument for us to accept it.

To sum up what we have so far, Joyce accepts, and so have we with him, that moral discourse is non-negotiably committed to the categorical *applicability* of moral obligations. Joyce wants to move from this to the claim that it is also a non-negotiable commitment of moral discourse that moral obligations are or entail *categorical reasons of rationality*. We have discussed Joyce's three arguments for that conclusion—the argument from Mackie's Platitude, the translation test argument and the use argument—in their various forms but we have found them wanting. So far we are not obliged to accept Joyce's claim that moral discourse wouldn't be *moral* discourse without categorical normative reasons.

Discussions of Joyce's attempt to settle the non-negotiable entailment claim in the literature occasionally attribute further arguments to him. One is the argument *from encouraging wrongdoing*.<sup>204</sup> According to it, when agents are aware of error theory and thus of their moral concepts not being imbued with practical clout, they can jiggle their desires round, thereby escape legitimate criticism of their behaviour, and thus feel free to perform wrongful actions. And because this is to be avoided (we don't want wrongs to be committed even more often than they already are) we should accept that moral reasoning comes with inescapable authority. However, this argument is clearly irrelevant on a reading according to which it is the *moral* wrongness that is doing the philosophical work in the argument. For that reading contradicts error theory by asserting that moral wrongness matters. But we interpret the ought as a prudential ought then the argument collapses into wishful thinking rather than an argument for the claim that practical clout is part of moral concepts.

Finally the literature also mentions the *redundancy of moral language* argument.<sup>205</sup> According to this argument, an alleged moral system without inescapable authority would make moral language redundant, but moral language isn't redundant, so it must contain some form of inescapable authority. This argument however clearly begs the questions against the naturalist who claims that,

---

<sup>204</sup> Tresan (2010: 234) based on Joyce (2006: 206).

<sup>205</sup> Tresan (2010: 235-6) based on Joyce (2006: 207).



strictly speaking, moral language is redundant. We need an *argument* that we can't have moral discourse without inescapable authority, not an *assertion* of that claim.

I will now leave Joyce's unsuccessful attempts to settle the non-negotiable commitment behind, assume for the sake of argument that he is right, and continue with his argument for the Substantive Claim that there are no categorical moral reasons of rationality.

### 2.3.2 Joyce's Substantive Claim

The claim that Joyce now has to argue against is the view he calls moral rationalism; viz., the view that moral reasons are reasons of rationality. As we know from the previous sub-section, according to Joyce, moral rationalism is the *only* way to make sense of the deeply embedded commitment in moral thought and talk to inescapable authority. For reasons of rationality are inescapably authoritative in the very same way that moral considerations are meant to be inescapably authoritative, and nothing else is. Unfortunately, argues Joyce, moral rationalism is false. Moral reasons are not a subset of normative reasons such that moral failing is, necessarily, rational failing.<sup>206</sup> This Substantive Claim, in combination with the conceptual claim that it is a non-negotiable commitment of morality that there are categorical moral reasons of rationality comprises Joyce's error theory.

Joyce's argument for the claim that there are no categorical normative moral reasons is that we can't "make sense" of such reasons.<sup>207</sup> We *can* make sense of reasons that "have some degree of dependence on [our] actual desires" but we cannot make sense of reasons that have no such dependence on our actual desires.<sup>208</sup> Practical reasoning "is *in general* a desire-sensitive affair, and ... thus so too are the practical reasons that come along with it".<sup>209</sup> Now we do have genuine practical reasoning as *some* desires can be shown to be misguided or inappropriate. This includes the desire to catch the 4pm train to Leeds if we know that we also desire to be at work in Leeds that day by 9am. For this is *desire-sensitive* practical reasoning: we can show that the desire to catch the 4pm train doesn't jibe with one's other desire to be at work on time. Of course this can't be the whole story—we'll also

---

<sup>206</sup> Joyce (2001: 80).

<sup>207</sup> Joyce (2001: 42; 2006: 190-9).

<sup>208</sup> Joyce (2006: 198).

<sup>209</sup> Joyce (2006: 198).

need some account of why the desire to be at work on time is the ‘more important’ one, and lots more besides—but the thought that Joyce is having is that we can at least get a grip on how practical reasoning with desires as input works. The problem for moral realism—whose fate, recall, is tied up with the fate of moral rationalism—is that turning practical reasoning into a desire-sensitive affair entails that not all desires can be shown to be misguided or inappropriate. This implies that “different people in the same circumstances may have very different reasons.”<sup>210</sup> So when the circumstances are ‘morally relevant’—i.e., if there is a question of killing or stealing—then one agent might have a reason to refrain from performing these actions, whereas another agent might have a reason to perform these actions, depending on what desires they have. But that means that not all reasons with recognisable moral content are categorical reasons—some will be desire-dependent or hypothetical. And that means that moral rationalism, and thus moral realism, is false.

I’ve just provided a very quick sketch of the intelligibility of practical reasoning as a desire-sensitive affair and I’ve asserted that according to Joyce, practical reasoning without desires as input won’t work. I will now expand on this second point. Joyce argues that practical reasoning without desires as input won’t work because we can’t allow agents to be ‘alienated’ from their reasons.<sup>211</sup> Joyce accepts that the presentation of a normative reason claim to an agent needn’t necessarily immediately motivate the agent. For Joyce accepts a *non-Humean* instrumentalist theory of reasons according to which some desires fall within the scope of rational influence.<sup>212</sup> Agents may need to generate desires that correspond with their reasons, and this process might not yet have affected the agent. But what Joyce doesn’t accept is that the agent, after having corrected his desires in light of full information about means-end relations in the instrumentalist way, can

---

<sup>210</sup> Joyce (2006: 198).

<sup>211</sup> Joyce (2001: 80).

<sup>212</sup> Joyce (2001: 69). The theory is non-Humean because Joyce accepts the interpretation of Hume’s theory of practical reasoning according to which no desires whatsoever fall within the scope of rational critique. The theory is still a version of instrumentalism “because it understands reasons only as means to an agent’s ends. An agent’s desires may be available for rational appraisal in a way with which Hume would have disagreed, and an agent may have X for an end (and thus a reason to pursue X) even though she does not desire X, but none of this distracts from the instrumentality of those reasons” (Joyce 2001: 77-8).

*legitimately* ask ‘but so what?’<sup>213</sup> For that would mean that if someone presents an agent with a contender reason claim then all that is going on is that this person is *barking a command* in the direction of the agent—a command, moreover, that doesn’t *resonate* with the agent. Such ‘barked commands’ are like commands to do a thousand push-ups even if there’s nothing about what you want to do that would make that command appropriate, legitimate or intelligible. Such demands aren’t reasons, thinks Joyce. Reasons ‘engage’ the agent in the right sort of way, for instance by being means to their ends. They certainly don’t *alienate* agents. And how do we avoid alienation? By “tying an agent’s normative reasons directly to the things that the agent is interested in”—that is, to his desires.<sup>214</sup>

Joyce thinks that the non-alienation constraint on reasons means that we get a dilemma for moral rationalism. Either we make sure that reasons *count as reasons* by tying them to agents’ desires because we thereby avoid alienation. But then we have to agree that reasons are relative in the sense that not all people might have a reason not to steal.<sup>215</sup> Or else we make sure that *all* people *do* have reasons not to steal, perhaps by prioritizing first-order moral reasoning and dropping the requirement that agents can’t be alienated from their reasons. But then we won’t have *reasons* because these ‘reasons’ aren’t tied in the right kind of way to our desires—they don’t *count* as reasons and are mere ‘barked commands’.

Even if we assume the non-alienation constraint, for the dilemma to work we still need to know, exactly, why rational deliberation without input from desires can’t generate categorical moral reasons of rationality. So far we have only seen a constraint on a successful attempt to argue that rational deliberation without input from desires can generate categorical moral reasons—the non-alienation constraint. But in fact this is all Joyce has. Joyce argues that *there is no reason to assume* that rational agents will converge in their judgments about which desires they ought and ought not to keep and thus places the onus of proof on the moral rationalist to show that when these agents think clearly and calmly about the matter, they will all desire not to steal.<sup>216</sup>

---

<sup>213</sup> Joyce (2001: 80-1).

<sup>214</sup> Joyce (2001: 81).

<sup>215</sup> Joyce (2001: 39, 41, 44).

<sup>216</sup> Joyce (2001: 76).

But how about idealization strategies? These are strategies that claim, roughly, that once agents reason in idealized circumstances, they *will* desire to do the morally right thing. In response to this claim in support of moral rationalism, Joyce insists that it matters *what* the idealization involves. Joyce argues that adding to the equation just that agents have all and only true beliefs makes no difference:

The differences that hold among actual agents that are based on doxastic disagreement will, of course, evaporate in their idealized counterparts, but there are no grounds for holding that all the other differences – those derived from different desires ... will likewise be lost in the idealization. So there will be no convergence and *a fortiori* no convergence regarding stealing ... etc.<sup>217</sup>

However, Joyce needs to do more than this, as various philosophers take up the challenge of showing why rationality demands moral behaviour. Joyce considers Smith's Kantianism.<sup>218</sup> Smith's argument that practical rationality is *substantively non-relative* (i.e., delivers reasons that aren't relative to people's idiosyncratic desires or wants) goes as follows. He asks us to consider "the empirical fact that moral argument tends to elicit the agreement of our fellows."<sup>219</sup> If that is true then it seems safe to say that if we think sufficiently long and hard about moral matters, we will agree on the same rules—that is, that we will all recognize the same reasons.

How about the rejoinder that this alleged 'empirical fact' is simply false? Smith replies by drawing attention to the following three alleged facts:

- 1        There are moral disagreements but there are also moral agreements
- 2        Of the moral agreements, many are due to rational deliberation
- 3        All of the disagreements are due to the failure of rational deliberation<sup>220</sup>

Joyce complains that 1-3 do not show that moral rationalism is true. 1 and 2 can be granted. Actually, 2 would be more interesting if it were formulated as 2\*:

- 2\*        All moral agreements are due to rational deliberation

---

<sup>217</sup> Joyce (2001: 76-7).

<sup>218</sup> Smith (1994); Joyce (2001: 88-91).

<sup>219</sup> Smith (1994: 188).

<sup>220</sup> Smith (1994: 188).

But Smith, wisely, doesn't profess on this issue. This is a wise decision because 2\* is more difficult to defend than 2, and because all the most important philosophical work in 1-3 is done by 3 anyway. After all, if 3 is true then someone holding the view that killing is permissible will be practically irrational. And since practical rationality generates reasons, if 3 is true then we all have a reason not to kill, even the person currently holding that killing is permissible. So if 3 is true then we get categorical moral reasons of rationality. But how do we argue for 3?

Smith employs the *Kantian Leap* to argue that the person wanting to kill—call him *the criminal*—holds an irrational desire. The Kantian Leap is the idea that when one deliberates about what one should do, one is in effect legislating for everyone. Everyone *in the same circumstances* as you would have the same reason that you decided you had. If the Kantian Leap is true then the idea is that if the criminal is rational, he wouldn't desire to kill—for by the Kantian Leap that would mean that we are all allowed to kill, and thus allowed to kill *the criminal*.<sup>221</sup> And no rational agent can allow for desires that potentially annihilate his own agency or capacity for rational deliberation.

But the Kantian leap is false, argues Joyce.<sup>222</sup> The problem lies with what we understand by 'same circumstances'. Either we understand 'same circumstances' to include the agents' idiosyncratic desires or we don't. If we do then it won't be true that we all have the same reasons. The criminal desires to kill and is willing to universalize that desire and take the risk of being killed—perhaps the criminal feels sufficiently safe with his guns and the support of his fellow thugs. I however don't to desire to kill—I don't want to take the risk of being killed myself—and so I want to universalize *that* desire. The problem for Smith, argues Joyce, is that neither is contradicting the other and that we'd need something *in addition to universalization* to settle the matter about who is irrational here—the criminal or myself. But the whole point about the universalization proposal was that *it* could settle this matter. Alternatively, if we exclude desires from 'same circumstances' then the Kantian Leap doesn't get us the claim that we all have reasons not to kill. This is because although we have a good conception of what it is for agents to universalize their

---

<sup>221</sup> Smith (1994: 195); Joyce (2001: 92).

<sup>222</sup> Joyce (2001: 92-4, 125).

actions based on their desires (the criminal vs. me-case), we can't make sense of the irrationality of desires that is wholly unconnected with what we presently desire.

Smith though offers the following thought to make sense of critique of an agent's desire in a way that doesn't involve pointing out a connection to other desires. The criminal

sticks with his opinion [that stealing is rationally acceptable] despite the fact that virtually everyone disagrees with him<sup>223</sup>

But of course, accepting arguments that most people accept is no virtue of rationality that we should aspire. Either way, therefore, the Kantian leap doesn't work.

Joyce's argument, it seems to me, works against Smith. The problem with Joyce's argument, though, is that it *only* works against Smith. Other Kantian strategies are still available. For instance, there are arguments to the extent that reasons can be Williams-internal (i.e., that something can be a reason if and only if it is suitably related to agents' antecedently held desires—accepted by Joyce) and yet categorical.<sup>224</sup> It would strengthen Joyce's case if we could do better than this.

Before closing I want to raise some further worries for Joyce's argument for the Substantive Claim that there are no categorical moral reasons of rationality. The first is that it is not at all clear that the non-alienation constraint on reasons is true. Do we really think that reasons can't be *reasons* once they fail to engage agents and what they care about, broadly construed? Indeed, don't we rather think that, as precisely our practice of moralizing shows, an agent can have a reason to do something (to give to famine relief for instance) *regardless* of whether he desires to kill? Furthermore, think about aesthetics and our general belief-forming processes. Don't we think that agents can have a reason (of rationality) to appreciate the aesthetic qualities of Rembrandt's *De Nachtwacht* even though they don't care about the great works of art? And don't we think that agents can have a reason (of rationality) to believe that the half-life of carbon-10 is 19.29 seconds when presented with the evidence even though they don't desire to engage with modern chemistry? Since all of these practices—the moral, the aesthetic and the

---

<sup>223</sup> Smith (1994: 94).

<sup>224</sup> Korsgaard (1986: 21). The reference to Williams is to his (1979).

epistemological—deal with categorical reasons and since this much, or so it seems anyway, is deeply ingrained in our ordinary thinking, Joyce’s move to wield in the non-alienation constraint to place into doubt especially moral reasons starts to look question-begging.

Second, if Joyce’s entire argument for moral error theory via the non-alienation constraint and his dilemma works, then it seems that Joyce is committed to the view that there are no categorical aesthetic and no categorical epistemic reasons of rationality either. That is an even tougher bullet to bite than just the bullet that there are no categorical moral reasons of rationality. Joyce however doesn’t consider this objection. In the terms of §1.4, Joyce accepts the (very implausible) first horn of the Formulation Dilemma, and this is a further strike against it.

The third worry is that Joyce accepts what he calls the ‘*practical reasoning theory*’ of reasons; the theory according to which our reasons are given by what agents would end up deciding to desire and do after flawless practical reasoning.<sup>225</sup> Copp professes that he finds it

counter-intuitive to suppose that the wrongness of torture depends on whether it is a necessary truth that anyone who believed that torture is wrong would be irrational to fail to take this into account in deciding whether to torture<sup>226</sup>

The fact that an account of moral wrongness strikes people who have thought hard and long about it as counter-intuitive is a further strike against it.

Fourthly, why can’t we argue that normative reasons are, to use one metaphor, *part of the fabric of the universe*, perhaps because reasons are facts and where ‘facts’ are understood here very broadly to cover how things are?<sup>227</sup> Perhaps the existence of such intrinsically normative facts *suffices* for the truth of the claim that we should all care about morality. This kind of view is what Korsgaard would count as a version of “normative realism” and provided the backbone for Mackie’s

---

<sup>225</sup> Joyce (2006: 194); also see Harman (1986: 66); Tresan (2010: 233n23).

<sup>226</sup> Copp (2010: 142-3).

<sup>227</sup> See Bedke (2012: 111) for this understanding of ‘facts’.

metaethics.<sup>228</sup> When defensible, the view would vindicate moral realism and a success theory of moral discourse.

Joyce doesn't consider this view. He would presumably reject it because it alienates agents from their reasons. Agents can ask 'so what?—what are these facts *to me*?' when they are presented with their existence. But at least offhand it seems that if there *are* reasons understood as intrinsically normative constituents of the world then, precisely, 'so what?' responses are not legitimate. If such facts 'emanate' categorical normative force, then, it seems, there is no question as to whether that force can be ignored.<sup>229</sup> The force is simply there, whether you want to recognize it or not. It might be difficult to believe that such (irreducibly) normative entities exist, but we'd need something like a queerness argument to show that they don't exist. And at least the queerness argument as we got it from Mackie was far from successful; moreover, Joyce doesn't have a queerness argument. In not considering this view Joyce is begging the question against it. Overall, then, Joyce's attempt to settle the Substantive Claim is problematic.

To sum up this sub-section, Joyce formulates a dilemma for the moral rationalist, whose theory he argues is the only hope for moral realism. Either reasons are *reasons* because they don't alienate people as they are sufficiently closely tied to agents' desires. But then we have to admit that different people might have different reasons in different circumstances, which is inconsistent with moral reasons' categoricity. Or else reasons are guaranteed to be the same for everyone because we deny that they require some kind of connection with our desires. But then these 'reasons' aren't in fact *reasons* as agents can then legitimately ignore them. I argued that the second horn of the dilemma is far from damaging for the moral rationalist and indeed the moral realist. For all that Joyce has, I explained, is some critique of existing Kantian strategies coupled with the remark that the burden of proof is on them to clinch an argument without violating the non-alienation constraint. And this leaves one escape route wide open for the success theorist: find a Kantian that can clinch the argument. I also argued that the non-alienation-constraint on reasons coupled with the practical reasoning theory of reasons can be seen as question-

---

<sup>228</sup> Korsgaard (1996: 39). Of course, Korsgaard argues that this view is false—but her arguments need not be accepted without a struggle, and other formulations of normative realism are possible. Also see Darwall *et al.* (1992: 118).

<sup>229</sup> Zangwill (2012: 345).



begging in the context of arguing for error theory, and some further objections beset his theory in addition to this. Moreover, Joyce does not have an acceptable answer to the Formulation Dilemma. So our conclusion has to be that Joyce hasn't made error theory sufficiently plausible.

However, again let's give Joyce the benefit of the doubt. Suppose his error theory can be established. Then what (prudentially) ought we to do with our error-ridden moral discourse?

### 2.3.3 Joyce's Solution

For Joyce the main consideration relevant to deciding what to do with moral discourse after error theory is what "is supported by some kind of cost-benefit analysis."<sup>230</sup> This includes having no replacement discourse at all, continuing as if we never discovered the truth of error theory, and everything in between. Joyce thinks that a fictionalist treatment of moral thought and talk does best on the cost-benefit analysis because it preserves important functions of our original tendency to moralize. In particular, fictionalizing bolsters self-control and combats weakness of will when it comes to performing those actions that we were used to call morally right and obligatory.<sup>231</sup> And, importantly, fictionalism procures these benefits without the great costs of having no replacement discourse at all (abolitionism) or violating the truth-norm that governs most of our utterances most of the time (preservatism).

Fictionalist theories of moral discourse either hold that we should believe fictionalist propositions like 'in the moral fiction, stealing is wrong' (content-fictionalism) or that we should *make-believe* or *pretend* that stealing is wrong (attitude-fictionalism).<sup>232</sup> Joyce is an attitude-fictionalist, and believes that if we *make-believe* moral propositions then, in every-day cases in which we aren't reminded of the heightened epistemic standards as exhibited in the philosophy seminar room, we will be more likely to refrain from stealing compared to when we just think about the prudential considerations that speak against stealing (such as

---

<sup>230</sup> Joyce (2007b: 68, also see 2001: 185).

<sup>231</sup> Joyce (2007b). Also see Olson (2011b: 182); Cuneo and Christy (2011: 93).

<sup>232</sup> Attitude-fictionalism is also sometimes called force-fictionalism (Enoch 2011a: 110n31).

increased possibility of ending up in jail).<sup>233</sup> Joyce believes that this is the case because ‘pretending’ that stealing is morally wrong isn’t some kind of ‘activity’ in the sense that in morally relevant situations we actively decide to pretend that there is a morality. Instead we bear an attitude of *precommitment* to morality.<sup>234</sup> Before we get ourselves into practically relevant circumstances we decide that we want to immerse ourselves in the moral fiction so that when it matters we simply and automatically assent to the proposition that stealing is wrong, and, especially when this is done vividly, we will have yielded motivationally efficacious emotions that make us refrain from stealing.<sup>235</sup> Only when we are in critical contexts in which we are most “undistracted, reflective, and critical” such as the philosophy seminar room do we forget about the fiction and agree that, really, nothing is morally right or wrong.<sup>236</sup> The reason why this works better than just thinking about the prudential benefits of not stealing is that humans have a tendency to rationalize changes in desires—‘I don’t really want to be a moral saint anyway, and no one’s looking, so stealing this sandwich is actually fine’.

Why can’t we simply go on believing in morality? That too would have by and large the same effect as *make-believing* in morality. Perhaps we can (largely) compartmentalize our (false) moral thoughts and our metaethical convictions.<sup>237</sup> Roughly, we should simply not think of error theory when it matters in daily life. Joyce objects that in doing this we will sacrifice the value of truth. Although truth isn’t intrinsically valuable, believing in falsehoods is a *practical* error because doing so is likely to have negative repercussions.<sup>238</sup> It is true that the norms of truth and truthfulness, which allow agents to retain their grip on reality, can be abandoned or overruled by norms of social usefulness. But they should not be overruled too

---

<sup>233</sup> Joyce (2001: 200).

<sup>234</sup> Joyce (2005: 305).

<sup>235</sup> Joyce (2001: 222-8).

<sup>236</sup> Joyce (2005: 290).

<sup>237</sup> Olson (2011b).

<sup>238</sup> Joyce (2001: 179). Similarly, Cuneo and Christy argue that “we should avoid making such errors because “all else being equal, having massive amounts of false beliefs is bad [so that] if someone were to find himself with packs of false beliefs about what Locke called ‘matters of maximal concernment,’ such as morality and religion, then he should want to remedy this” (2011: 95)

often.<sup>239</sup> The bad consequences of this are “doxastic schizophrenia ... which can ... be expected to lead to various kinds of pragmatic handicap.”<sup>240</sup> Another bad consequence is that a seemingly useful false belief will require all manner of compensatory false beliefs to make one’s entire package of beliefs coherent or internally consistent.<sup>241</sup> This is not feasible.

So perhaps Joyce is right about the benefits of fictionalism over having no replacement discourse and preserving our error-riddled discourse, *if it works*. But can fictionalism be made to work on its own terms? That is, is it true that we can reap the benefits of social cooperation with a fictionalist system that is supposed to regulate our practical deliberations, where this system is not undergirded by real normativity?

One point of critique comes from Lillehammer, who asks how we can decide which values we are to further with the fictionalist discourse without thereby letting *moral* reasoning in through the back door again.<sup>242</sup> Joyce argues that we have to look at what is “instrumentally useful”.<sup>243</sup> And what is instrumentally useful is in part determined by the kinds of creatures we are.<sup>244</sup> For us, cooperation is generally useful, and so it would be useful to adopt the policy of immersing ourselves in the moral fiction.<sup>245</sup> The question though is what happens when lots of people disagree about what is instrumentally useful. What do we do with people who’d rather see very little cooperation because that benefits their business? How do we decide what becomes part of the moral fiction, given that, as agreed, there are no *moral* rules constraining that content? Just looking at the kinds of creatures we are might not work. Some people are very different from others.

---

<sup>239</sup> Lillehammer (2004: 105) calls this a “sane commitment” to truth, contrasting it with dogmatically believing that the norms of truth and truthfulness are to be upheld all of the times.

<sup>240</sup> Joyce (2007b: 68).

<sup>241</sup> Joyce (2001: 179).

<sup>242</sup> Lillehammer (2004: 103).

<sup>243</sup> Joyce (2001: 184).

<sup>244</sup> Joyce (2001: 182).

<sup>245</sup> Joyce (2001: 212).

A further worry is whether Joyce can “dispel [the worry] that making fictive judgments need involve self-deception or irrationality”.<sup>246</sup> Joyce argues that self-deception is avoided because it is always open to the make-believer to opt out of the fiction and agree, in the philosophy seminar room perhaps, that in reality morality is a sham.<sup>247</sup> But this creates a tension with the practical benefits of fictionalizing. Once we are very much engrossed in the moral game, it seems, we will be in a state of mind that excludes thinking about error theory and this may indeed get us to refrain from stealing. But if this happens it also seems as though we are *deluding* ourselves: by hypothesis, the thought that error theory is true isn’t accessible if we are very much engrossed in the moral game. For if it were then the practical benefit of setting the agent on a course to not stealing wouldn’t occur. However, once we try to avoid this from happening by allowing ourselves to become conscious of the thought that error theory is true when we consider to steal something, then, especially when we know that no one is looking, the practical benefits of fictionalism will be compromised. So it is not clear that Joyce can avoid the worry of self-deception whilst retaining the benefits of some form of moral thinking. Indeed, it seems that we have a dilemma. Either accept self-deception but that sounds just as bad as compromising on truth. Or avoid self-deception, but then the practical benefits of fictionalizing will be compromised. Fictionalism is in trouble.

This was a short sketch of fictionalism because I will return to it specifically in Chapter 8. Nonetheless we can already conclude that it is not free of problems. Overall, we should conclude that Joyce’s error theory is not plausible. The theory solves none of the four major problems for error theory as discussed in §1.4.

## 2.4 Other Error Theories

In this section I provide a briefer discussion of the commitments of other existing error theories. Again my aim is to show that the four problems from §1.4 have not been adequately dealt with by error theorists.

I start with Olson’s error theory. Olson stays fairly close to the error theories of Mackie and Joyce. He accepts what he calls the standard formulation of error theory:

---

<sup>246</sup> Joyce (2001: 194).

<sup>247</sup> Joyce (2001: 196).

**Standard Formulation**

Error theory concerns ordinary folk moral discourse, the mode of commitment of moral discourse to the poisoned thesis is one of conceptual entailment, the content of the poisoned thesis is that there exist categorical moral reasons of rationality, and moral statements are false<sup>248</sup>

Moral facts entail facts about categorical reasons, but there are no categorical reasons (these are too queer to exist) and hence, by entailment, there are no moral facts.

Like Mackie, Olson thinks that the normativity that attaches to hypothetical reasons can be reduced to naturalistically respectable elements:

The fact that there is (conclusive) non-categorical reason for chess players not to move the rook diagonally just is the fact that moving the rook diagonally is incorrect according to the rules of chess [there is nothing queer about this and likewise] there is nothing metaphysically queer about the fact that there is non-categorical reason for a soldier to comply with the orders of a general, since this is just the fact that complying with the orders of those superior in military rank is part of the role of being a soldier<sup>249</sup>

The reason why there is no queerness here is—focussing on the chess example for simplicity—that these norms are *constitutive* norms, determining simply *what it is* to play the game.<sup>250</sup> Had we been talking about the *substantive* norms of playing chess, such as one obliging us *to play*, then this move would not have been possible. Moral norms are substantive and don't exist; the norms of chess aren't substantive and exist.

Regarding the question of what to do with moral discourse, Olson is a *preservatist*, thinking that we can continue to indulge in our error-ridden moral

---

<sup>248</sup> Olson (2011a: 62). Recently however Olson had changed his mind about the 'content of the poisoned thesis': "In previous work, I maintained that moral facts are queer because they are or entail categorical reasons (Olson 2010, 2011). I now believe that the best articulation is that moral facts are queer in that they are or entail facts that count in favour of certain courses of behaviour, where the favouring-relation is irreducibly normative" (forthcoming-b: 13).

<sup>249</sup> Olson (2011a: 65); also see (2011b). Bedke similarly writes "what it is for F to be a reason for A to  $\phi$  is for A to have a goal with content F and for  $\phi$ ing to further that goal" (2010: 50).

<sup>250</sup> Also see Evans and Shah (2012: 83n7).

discourse as long as we *compartmentalize* our moral and metaethical thoughts.<sup>251</sup> The thesis that we screen off metaethical considerations from moral considerations when we are (a) error theorists and (b) can stand a lot to win by displaying selfish behaviour is extremely implausible, at least phenomenologically, but also philosophically as I will demonstrate in Chapter 8.

The fact that Olson's error theory is so similar to that of Mackie and Joyce suggests that it suffers from the same kinds of problems as theirs. And indeed it does. First, for Olson just as much as for Mackie we seem to have not a reduction but an elimination of normativity. Using our conceptualization of reduction relations and the way in which they can be more or less reductionist 'in substance' from the previous section, it could be argued that Olson's alleged reduction violates too many of our commonsensical claims about normativity and therefore counts as an elimination of normativity. Furthermore—and this is the second point—this would open up Olson to the further objection that his local error theory or moral discourse collapses, implausibly, into an error theory of normativity as such. Finally, as just indicated, Olson doesn't seem to have an adequate answer to the question of what we ought to do with moral discourse after error theory.

The next error theory I discuss is Schiffer's. Contrary to Olson, Schiffer argues for a very different error theory from the error theories discussed so far. According to Schiffer, moral properties do exist albeit only in the sense of being *mere shadows of predicates*.<sup>252</sup> The idea is that every meaningful predicate F has a nominalization, 'the property of being F', which cannot fail to refer. The property of being F is a pleonastic gift of the predicate F, but this does not entail that moral properties are also *instantiated* in a particular world. For being a property understood as a pleonastic shadow of a predicate is one thing and pleonastically easy to come by; being a property that is actually present in a given possible world is quite another thing. Unfortunately, argues Schiffer, moral properties cannot possibly be instantiated because they don't have instantiation conditions at all.<sup>253</sup> That moral properties don't have instantiation conditions falls out of Schiffer's *anti-Platonist*,

---

<sup>251</sup> Olson (2011b).

<sup>252</sup> Armstrong (1989: 78).

<sup>253</sup> Schiffer (1990: 603-4, 606-7); also see his (2003: Chapter 6). Lillehammer (2004) and Loeb (2008) make similar points.

*'use' theory of sameness of meaning*. Platonists in the relevant sense hold that two predicates have the same meaning if and only if the property that gives the meaning to the predicates can be identified and if and only if both predicates are relevantly semantically related to that property. Schiffer's use theorist holds that sameness of meaning can be established by showing relevant sameness of *use*, where this issue is specifiable without mention of an expression's being related to a property. Unfortunately, given the persistence and omnipresence of 'faultless' moral disagreement—disagreement that obtains in the absence of false non-moral beliefs and the like—we can't establish relevant sameness of use. There is no moral proposition that is relevantly determinately true. People's varying uses of moral terms can't supply (determinate) instantiation conditions for moral properties. All we can say is that two parties have different (ultimate) moral principles and hence different criteria of application for their moral terms.

Schiffer does think that moral reasoning and language has an *instrumental* role to play in our lives even though we should accept his error theory. For we can best fulfil our desire to live in what "one would have called the morally best world" by *using* moral talk without assenting to any of its propositions (though Schiffer leaves it open what the nature of this 'use' is).<sup>254</sup> This comes close to Mackie's revolutionary cognitivism.

Schiffer's version of error theory faces problems. First, it requires error theorists to accept controversial theories about the nature of moral properties and moral semantics. This is not a knock-down argument against this way of formulating error theory, but it would be preferable if error theorists could formulate a theory that doesn't suffer from these problems. Second, Schiffer's solution to the problem of what we ought to do with moral discourse once we become convinced of error theory is not worked out to a degree of sufficient detail. This is also not a knock-down argument against his error theory, but it would be better if we had more to say (I will try to say more about this kind of position in the final chapter of this thesis). In general, although Schiffer's error theory is very interesting, it is too far removed from the kind of Mackie-Joyce-style error theory I want to resuscitate in this thesis. I will therefore put it aside from now on.

---

<sup>254</sup> Schiffer (1990: 614).

I close with Streumer's version of error theory, which is currently defended in various complementary papers. One paper deals with the question 'do normative judgments aim to represent the world?'<sup>255</sup> Streumer answers this question in the affirmative. He also argues that this entails that normative judgments are beliefs that ascribe normative properties. Here the term 'property' is used in a non-minimalist sense—i.e., not in Schiffer's sense in which properties are simply the 'shadows' of predicates. Hence Streumer argues for a global error theory about all normative judgments, including moral and epistemic ones. Streumer also argues that there are no normative properties.<sup>256</sup> Suppose normative properties are identical to descriptive properties. This entails, argues Streumer, that we should be able to say which descriptive properties normative properties are identical to. But we aren't able to say that. Streumer attempts to show this with several arguments, one of which resembles Schiffer's argument from faultless disagreement, although Streumer doesn't mention Schiffer. So normative properties aren't identical to descriptive properties. Moreover, there are no *irreducibly normative* properties either.<sup>257</sup> This entails that there are no normative properties at all. Importantly, we reach this argument without having said anything about the thorny issue of whether normative properties can be descriptive properties.<sup>258</sup>

Streumer's other paper that is directly relevant to error theory argues that we cannot believe error theory and that this is good news for error theory.<sup>259</sup> We cannot believe error theory because error theory entails that there is no reason to believe error theory. For reasons require normative properties and there are no normative properties. But, argues Streumer, we cannot fail to believe what we fully believe to be entailed by our beliefs. This means that when we believe error theory we cannot but believe that there is a reason to believe the error theory. Unfortunately, it is true of beliefs that we cannot have them without believing that there is a reason for that belief. This entails that we cannot believe the error theory. Moreover, given that there cannot be a reason for someone to believe anything (including error theory) if

---

<sup>255</sup> Streumer (2008).

<sup>256</sup> Streumer (2011).

<sup>257</sup> This is because, according to Streumer, Jackson's (1998) argument for that claim works; Streumer (2008, 2011: 338-9).

<sup>258</sup> Streumer (2011: 346).

<sup>259</sup> Streumer (forthcoming).



this person cannot believe that thing, there is no reason to believe error theory either.<sup>260</sup>

Why is this good news? First of all, it is not a *problem* that we cannot believe the error theory. It is clearly not a problem for any theory if we do not believe it, so it seems that it also shouldn't be a problem for a theory if we *cannot* believe it. Second, we can reject:

### **Objection from Toothlessness**

If there is no reason to believe error theory then error theory is polemically toothless. For if there is no reason to believe error theory then it is not a rational mistake to reject error theory<sup>261</sup>

We can reject Objection from Toothlessness because we can believe the various *parts* of error theory at different times and that these parts *seem* to show that error theory is true, in which case—on the assumption that we rationally should believe what we have most evidence to believe—it *is* a rational mistake not to assent to these seemings.

The second benefit of the fact that we cannot believe error theory, on Streumer's view, is that we no longer need an answer to the question what we ought to do with moral discourse once we become convinced that we should fully believe error theory. For we cannot fully believe error theory.

However, there are problems with Streumer's error theory. The first of these is that Streumer explicitly accepts a global error theory of all normative judgments, including moral and epistemic ones. True, Streumer does argue that this generates benefits for his view, but in fact this can be questioned. For it is really hard to believe that there are no reasons for beliefs. Another problem comes from the fact that we can't believe error theory in its entirety but that we *can* believe its various parts, for this entails a schizophrenic attitude towards morality. We can't fully believe error theory but we know that its various parts are true (and together they entail error theory). This also in fact, and contrary to what Streumer claims, does make the question of what we ought to do with our error-riddled moral discourse

---

<sup>260</sup> Also see Hampton (1998); Shafer-Landau (2003: 205-6); Olson (2009a).

<sup>261</sup> Cuneo (2007a: 117-8).

relevant. For in practical situations it is all too easy to believe that there is something fundamentally wrong with morality even if we cannot fully believe error theory, in which case the natural bulwark that morality provides—in the form of a believe-led inhibition against stealing—may disappear. But then we need a theory about whether we can have a surrogate bulwark that isn't *morality* (which is in error), or else we need an argument why we can still use morality as a bulwark even though it is in error, or that we can do without morality or something like it. The problem is that Streumer doesn't consider these issues.

Overall we reach the following conclusion. There are error theories that are relevant similar to that of Mackie and Joyce and there are also error theories that are different. From the latter category I discussed Schiffer's theory. Although this error theory is interesting, I argued, it is also not obvious that it will work, and, in any case, my aim here is to revive the Mackie-Joyce-style error theory, so I put it aside in what follows. Error theories that are relevantly similar to that of Mackie and Joyce (Olson's and Streumer's) suffer from similar problems and have room for improvement, which I will aim to provide in the remainder of this thesis.

## **2.5 Conclusion and Preview**

In this chapter I have discussed the main error theories currently on offer; I discussed the theories of Mackie and Joyce in significant detail and the theories of Olson, Schiffer and Streumer more quickly. I argued that none of these theories can solve the four main problems for error theory as identified in §1.4. The remainder of the thesis aims to do better. To this end it starts, in Chapter 3, with an investigation of the options for error theorists when it comes to formulating their theory.

## Chapter 3

# Options for Error Theorists

### 3.1 Introduction

In this chapter I formulate a direct and maximally non-committal formulation of error theory that acts as a template for specific, committal formulations of error theory (§3.1). The idea is to canvass what the options for error theorists are so that a possibly more plausible formulation of error theory can be defended than has so far seemed possible. The generic formulation of error theory that I will formulate has three places where theoretical commitments are required to get the argument to work, and the chapter subsequently discusses what the possibilities are (§3.2-4). The chapter also discusses the different answers to the following question: what (non-morally) ought we to do with our error-riddled moral discourse (§3.5)? It closes with a conclusion and preview (§3.6).

In addition to using a direct argument for moral error theory based on the template that I am about to present it is also possible to formulate indirect arguments for error theory. Before presenting my template I will explain why these alternative ways of defending error theory are not very useful. One possibility is to formulate an argument from an *inference to the best explanation*.<sup>262</sup> According to that argument, an inquiry into the past 2000 years of metaethical endeavours justifies claiming that the best explanation for our failed attempts to vindicate moral realism is that moral error theory is true. A second possibility is to formulate an *argument from elimination*, which relies on direct arguments against every rival of the moral error theory, and concludes, by elimination, that moral error theory must be true.<sup>263</sup> Both arguments leave open the possibility that, in fact, moral error theory is false. A better, because more direct, argument for error theory goes as follows (cf. §1.1 of this dissertation):

---

<sup>262</sup> Joyce (2007b: 52); Burgess (2010: 13).

<sup>263</sup> Joyce (2001: 101).

### **Moral Error Theory**

#### **P1 Non-Negotiable Commitment Claim:**

Moral statements and judgments carry a *non-negotiable commitment* to a robustly understood moral reality; the moral reality thesis

#### **P2 Substantive Claim:**

This commitment is flawed

#### **P3 Auxiliary Claim:**

If moral discourse carries a false non-negotiable commitment then it is itself flawed

#### **C Conclusion**

Moral discourse is flawed: the statements and judgments comprising it are 'in error' [From P1, P2, P3]

The three choice points that this template provides are the phrase 'is non-negotiably committed', which concerns the nature of moral discourse's mode of commitment to T, the content of 'the moral reality thesis', and 'flawed'. I will now consider the contender interpretations for these placeholders (§3.2-4).

## **3.2 The Non-Negotiable Commitment Claim**

This section discusses the various options for understanding the nature of moral discourse's mode of commitment to the moral reality thesis. Because moral discourse is comprised of moral statements of various kinds we can use the familiar options from the philosophy of language for understanding how such linguistic utterances can carry semantic or pragmatic commitments, which are, in the relevant sense and to varying degrees, non-negotiable.

In previous chapters I have already introduced two of the relevant options; viz., conceptual and metaphysical entailment. In that context, which was concerned with understanding conceptual and metaphysical entailment well enough to be able to appreciate some problems with Mackie's and Joyce's theories, it wasn't necessary to consider these options in much detail. Here I will provide more detail, starting with a fuller investigation of the nature of the various moral statements that comprise moral discourse.

Consider:

<b>Wrong</b>	Stealing is morally wrong
--------------	---------------------------

Wrong is an atomic moral statement; ‘atomic’ because it has the logical form ‘ $x$  is  $F$ ’, and ‘moral’ because the predicate  $F$  is a moral predicate. It is possible to embed Wrong under various linguistic operators, such as negation, modal and question operators. Doing that gives us the following non-atomic moral statements:

<b>Negation</b>	Stealing is not morally wrong <sup>264</sup>
<b>Modal</b>	Stealing might be morally wrong
<b>Question</b>	Is stealing morally wrong?

It is natural to assume that only atomic moral statements are committed to the moral reality thesis, as only they are, at least *prima facie*, in the business of ascribing moral properties to action types and tokens (and character traits, states of affairs, *et cetera*). But in fact some of the other moral statements can also somewhat naturally be interpreted to commit its utterer to the moral reality thesis. For instance, a denial of the wrongness of stealing is not plausibly heard as committing its utterer to the wrongness of stealing (to the contrary). However, this leaves it open that Negation carries a commitment to the existence of a moral reality in a different way. Perhaps if stealing is not wrong it is *positively morally permissible*; i.e., permissible according to morality. I discuss this and other complexities in the next chapter. For ease of exposition I will here only have atomic moral statements in mind when I speak of ‘moral statements’ and their mode of commitment to the moral reality thesis.

### 3.2.1 Entailment

One option is that ‘is non-negotiably committed to’ is best understood as ‘entails’. I make three explanatory remarks about entailment relations in general before discussing the different kinds of entailment relations that are options for error theorists. First, entailment relations primarily have propositions as their relata and

---

<sup>264</sup> To avoid unnecessarily longwinded formulations I take Negation to be equivalent to ‘it is not the case that stealing is morally wrong’. I also use the negation in Negation in its ordinary truth-functional rather than metalinguistic way.

are most fruitfully understood in terms of truth-conditions.<sup>265</sup> If a proposition  $p$  entails proposition  $q$  then  $p$  can't be true without  $q$  being true.

A second explanatory remark is that entailment relations hold regardless of whether anyone recognizes them.<sup>266</sup> For instance, a certain answer could be entailed by the way in which a particular mathematical question is phrased even if no human being (or rational agent for that matter) has so far realized this. This means that entailment relations can hold between moral statements even whilst the folk, or the philosophers for that matter, are not aware of this.

A third and final remark is about what it is in virtue of which the entailment relation obtains. This can be either in virtue of logic, concepts, or metaphysics. Hence there are three distinct kinds of entailment relations; viz., logical, conceptual and metaphysical entailments. I discuss each of them below. I explain how the entailment works in standard cases and what must be true for the kind of entailment under discussion to work in the moral case.

### 3.2.1.1 Logical entailment

If proposition  $p$  logically entails proposition  $q$  then  $q$  is true whenever  $p$  is true and in virtue of a logical connection between  $p$  and  $q$ . A standard case is provided by the (conjunctive) proposition 'all Fs are Gs and this is an F' ( $p$ ), and 'this F is a G' ( $q$ ).<sup>267</sup> The latter is logically entailed by the former because it is logically impossible for it to be true that all Fs are Gs and for this F not to be a G.

Logical entailment is a very unlikely candidate for the kind of non-negotiable commitment relation that obtains between moral discourse and the moral reality thesis (if indeed there is such a relation). This is because a further feature of logical entailments is that they are content-*independent*, whereas what we have in the case

---

<sup>265</sup> Streumer (2007: 355) and Copp (2007: 130-1) respectively. On a stretch of terminology, entailment relations can also be thought to hold between facts. I won't take this option into consideration here. Also see Soames (2006).

<sup>266</sup> Street (2010: 367).

<sup>267</sup> For ease of exposition I use a conjunctive proposition (because that allows me to talk about just two propositions). If you don't think that it is useful to talk about conjunctive propositions (because you might think that they are not really propositions but just two separate propositions), then you can read me as saying that logical entailments hold between *sets* of propositions, e.g., between  $p$  ('All whales are mammals'),  $q$  ('This is a whale'), and  $r$  ('This whale is a mammal'). See Anscombe (1959: 32).

of moral discourse and the moral reality thesis is a content-*dependent* entailment, if indeed an entailment relation is relevant for understanding its commitments. Therefore, I set logical entailment aside in what follows.

### 3.2.1.2 Conceptual entailment

Conceptual entailments are similar to logical entailments in that they are best understood in terms of propositions and truth-conditions: if a proposition *p* entails proposition *q* then *p* can't be true without *q* being true. However, truth is here preserved not as a matter of content-independent logical necessity but in virtue of content-dependent connections between concepts (they may disclose analytic truths).<sup>268</sup> As I explained in my discussion of the error theories of Mackie and Joyce, this is currently the most popular option, and amounts to:

#### **Conceptual Entailment Claim**

Moral discourse's mode of commitment to the moral reality thesis is one of conceptual entailment

The Conceptual Entailment Claim has thus far been the dominant theory for thinking about the mode of commitment of moral discourse to the moral reality thesis.<sup>269</sup> It can be made more precise in at least the following two ways.

One way of explicating the notion of conceptual is to formulate an *epistemic notion of conceptual entailment*, according to which a proposition is a conceptual truth just in case any person who meets the minimal conditions required for grasping it would thereby be in a position to know that it is true merely by considering carefully whether it is true while thinking clearly.<sup>270</sup> For instance, 'all vixens are female foxes' is a conceptual truth because anyone can come to know that it is true merely by reflecting carefully on what it says (so long as they are minimally competent at grasping it). Likewise for propositions with moral content, says this

---

<sup>268</sup> A proposition is analytically true, if it is, solely in virtue of its meaning (Juhl and Loomis 2010: ix-x; Williamson 2007: 60).

<sup>269</sup> The Non-Negotiable Commitment Claim is often simply called the *Conceptual Claim* (Smith 1994: 65; Joyce 2007b: §4; Robertson 2008: 111; Joyce 2001: 5; Shafer-Landau 2005: 108; Brink 1984; Copp 2010; Tresan 2010).

<sup>270</sup> Copp (2007: 121). Williamson similarly writes about the epistemological conception of analytic truths such as 'Every vixen is a female fox' that "failure to assent [to such truths] is not merely *good evidence* of failure to understand; it is *constitutive* of such failure" (2007: 73).

proposal. 'If stealing is morally wrong then there exist Mackie-Objective and Mackie-Prescriptive moral properties' would be a conceptual truth, and would tell us what has to be true for 'stealing is morally wrong' to be true, solely in virtue of it being the case that agents who meet the minimal conditions required for grasping this proposition would be in a position to know that it is true in the way suggested.

This way of explicating the idea of conceptual entailment faces the following difficulty. It requires us to specify in a non-question begging way what it takes for an agent to be minimally conceptually competent, but we better not argue that minimally conceptually competent agents are those that regard the proposition 'all vixens are female foxes' as conceptually true. That would be question-begging.

Another way of spelling out the notion of conceptual entailment is to focus directly on the *applicability conditions* of the concepts involved. Take moral concepts and the atomic moral statements that record their instantiation. Moral statements entail statements about categorical moral reasons (say) because it is necessary to the correct applicability of moral concepts, as expressed by moral statements, that there exists categorical moral normativity.<sup>271</sup>

This is the explication of conceptual entailment that I used to formalize Mackie's and Joyce's account of the conceptual entailment claim earlier in the thesis. The sentence 'this is a vixen' entails 'this is a fox' in virtue of the concept VIXEN expressed by the term 'vixen' in the sentence 'this is a vixen'. For the applicability conditions of vixen are two-fold: in order to use vixen correctly, you have to apply it to something that is both female and a fox. Take either femaleness or foxhood out of vixen and the *schvixens* you end up referring to simply don't deserve the label 'vixen'. Likewise, 'stealing is morally wrong' entails 'there exist objective and prescriptive values' in virtue of the concept MORALLY WRONG expressed by the term 'morally wrong' in 'stealing is morally wrong'. For the applicability conditions of morally wrong are two-fold: in order to use morally wrong correctly you have to apply it to something that both instantiates an objectively prescriptive value and that is of recognisable moral import. Take

---

<sup>271</sup> Also see Finlay (2008a: 365).



categorical reason-giving force out of moral concepts, and the *schmoral* concepts you end up with are too “wimpified” to be mistaken for the real thing.<sup>272</sup>

One issue that arises with regard to this is how, exactly, we are to understand the nature of the *correctness* of the application conditions. For ‘correct application conditions’ is ambiguous between ‘conditions of true application’ and ‘conditions which constrain how competent concept-users apply the concepts’ and perhaps other readings as well. This makes it difficult to specify what these conditions are in a non-question-begging way. Moreover, we have already seen problems with Mackie’s and Joyce’s attempts to make the Conceptual Entailment Claim work. Nevertheless, contrary to logical entailment, conceptual entailment isn’t obviously useless for error theory, so we should keep it as an option for error theorists.

### 3.2.1.3 Metaphysical entailment

Metaphysical entailments are similar to logical and conceptual entailments in that they are best understood in terms of propositions and truth-conditions: if a proposition *p* entails proposition *q* then *p* can’t be true without *q* being true. However, truth is preserved not as a matter of logical or conceptual necessity but in virtue of a metaphysical or substantive necessity “of the same modality as the connection between being gold and having atomic number 79”.<sup>273</sup> For another example: the proposition ‘this is water’ is metaphysically entailed by the proposition ‘this is H<sub>2</sub>O’ because it is a metaphysical necessity that water is H<sub>2</sub>O—there is no possible world in which something is water but not H<sub>2</sub>O.<sup>274</sup>

Importantly, conceptual and metaphysical entailment relations differ only about what it is in virtue of which the entailment holds (this yields concomitant differences in the epistemology of both entailment relations). But since both are *entailment* relations this difference doesn’t compromise the ‘stringency’ of either relation. They are equally necessary relations; the point is just that their necessity can be explained by reference to different things. So both can be used as models for the error theorist’s Non-Negotiable Commitment Claim. You can’t have morality without objective prescriptivity either because denying that we need objective

---

<sup>272</sup> Joyce (MS).

<sup>273</sup> Railton (1997: 70).

<sup>274</sup> Putnam (1975); Kripke (1972). Cf. Farrell (1981). Also see Shafer-Landau (2003: 84); Fine (2005); Vaidya (2010).

prescriptivity for morality exhibits a conceptual mistake or because it displays a more general mistake about the nature of morality without that mistake implying anything about one's mastery of the concepts involved.<sup>275</sup>

Just as with conceptual entailments, there are various ways of establishing metaphysical entailment relations. I mention two of them here.

The first strategy is an indirect strategy. Error theorists can critically assess the proposals of such synthetic naturalists as Brink, Boyd and Railton and argue that their proposals about the nature of moral rightness and wrongness (goodness, badness, obligatoriness, *et cetera*) are fundamentally lacking in those respects that make error theory plausible. Synthetic naturalists typically claim that the nature of moral goodness consists in some complex natural property, and that the fact that this property is intimately connected to a person's (distinctively) human capacities, including that person's well-being, explains and grounds that natural property's normative role.<sup>276</sup>

By critically assessing this kind of moral naturalism error theorist can reach a metaphysical entailment claim. So error theorists may argue that the normative role that the natural properties play in the synthetic naturalists' theory in fact aren't sufficiently 'tight' or 'internal' to account for what we generally or even platitudinously consider moral properties' normative role to be. This gives rise to a metaphysical entailment claim because the considerations that are brought to bear by the error theorist in providing an account of the nature of moral goodness and wrongness *et cetera* can be *a posteriori*, *non-conceptual* considerations which nevertheless specify what it takes for moral goodness to *be* moral goodness.

The second strategy available to error theorists for arriving at a metaphysical entailment claim is more direct. On this strategy, error theorists muster intuitions, thought-experiments and platitudes about the nature of moral wrongness—derivable from reflections on the workings of moral discourse, and therefore *a posteriori*—and argue that the best explanation of these platitudes *et cetera* is that there is an entailment relation between the proposition that X is morally wrong and that of X being authoritatively or categorically required by practical rationality (or some

---

<sup>275</sup> For similar points see Fine (2002: 255).

<sup>276</sup> Brink (1984, 1989); Boyd (1988); Railton (1986, 1989, 2003).

such).<sup>277</sup> Again, there is an entailment relation but not one in virtue of connections between concepts but instead in virtue of *a posteriori* considerations about the nature of moral wrongness itself.

### 3.2.2 Implicature

A second possibility is to understand the locution ‘is non-negotiably committed to’ as a placeholder for ‘implicates’. Implicature relations hold just in case a *statement* *q* (as opposed to *proposition* *q*) is linked to another *statement* *p* without *q* being part of the truth-conditions of *p*.<sup>278</sup> Implicatures come in two kinds: conversational and conventional.<sup>279</sup> For both kinds, I explain how the proposal works in standard cases, and then I explain what must be true for it to work in the moral case.

#### 3.2.2.1 Conversational Implicature

Normally conversational implicatures hold between statements as they are uttered by speakers in particular contexts, not between statements *per se*, independently of the context of utterance.<sup>280</sup> Conversational implicatures are standardly understood along Gricean lines. Grice explains implicatures and how they arise with the aid of two phenomena: the cooperative principle and conversational maxims.

The cooperative principle says:

[m]ake your conversational contribution such as is required at the stage at which it occurs, by the accepted purposes or direction of the talk exchange in which you are engaged<sup>281</sup>

The maxims tell you what to do to meet the principle; e.g., the maxim of quantity tells you to make your contribution neither less nor more informative than is required by your talk exchange, and the maxim of relation tells you to make your contribution relevant. According to Grice, we are justified in assuming that people conform to the cooperative principle and hence to the maxims (absent special

---

<sup>277</sup> Further theoretical considerations could be added to these intuitions, platitudes and thought-experiments; e.g., Enoch’s deliberative indispensability argument (2011a: Chapter 3). Also see Scanlon (2009, Lecture 2: 12).

<sup>278</sup> Finlay (2005: 1); Wayne (2008).

<sup>279</sup> Finlay (2005: 1-2).

<sup>280</sup> ‘Normally’, because I will discuss an exception shortly.

<sup>281</sup> Grice (1989: 6).

circumstances such as when people are acting). So for example, if you ask me ‘Will Pete be at the meeting tomorrow?’ and I answer ‘his car broke down’ then my utterance implicates that he won’t be at the meeting. For the implicature is *calculable*: you make the justified inference that my original statement has this implicature because that is the best way you have of making sense of my obligation to abide by the maxim of quantity as a means of abiding by the cooperative principle. After all, if my statement didn’t carry this implicature then I would be saying something less informative than was required to answer your question whether Pete will be at the meeting.<sup>282</sup>

Note that the implicature is also *cancellable*: it is a felicitous addition to my statement ‘his car broke down’, and ultimately to our conversation, if I continue to say ‘but mind you, Pete will be at the meeting tomorrow—he’ll go by bike.’<sup>283</sup> I myself might go through your calculation of the implicature, realise that you justifiably make the inference that my statement implicates that Pete won’t be at the meeting, and hence make this implicature-cancelling addition to save the conversation.

The fact that conversational implicatures are cancellable means that they cannot be used in a defence of moral error theory. For moral error theory requires a *non-negotiable* commitment and cancellable commitments clearly do not count as non-negotiable in the relevant sense. So we can safely put this option aside.

### 3.2.2.2 Conventional Implicature

Conventional implicatures also hold between statements rather than propositions, but they do so in virtue of the (linguistic) meaning of one or more words that feature in the relevant (target) statements. This sets them apart from conversational implicatures, which hold in virtue of the cooperative principle and the maxims.<sup>284</sup>

Here is an example:

---

<sup>282</sup> This is an example of a *particularized conversational implicature*, as the utterance in the text does not standardly carry said implication and is dependent on features of the context to arise (if utterances standardly carry implications and are not dependent on conversational context we speak of *generalized conversational implicatures*; Strandberg 2012: 93). The latter are nonetheless implicatures because they also depend on the cooperative principle and conversational maxims to arise.

<sup>283</sup> Copp (2007: 172).

<sup>284</sup> Grice (1975); also see Davis (2010).

Target statement:	‘Thora is a baby but she is quiet’
Descriptive meaning:	‘Thora is a baby and she is quiet’
Conventional implicature:	‘Babies are not usually quiet’ <sup>285</sup>

This is an instance of conventional *implicature* and not entailment because the sentence ‘Thora is a baby, but she is quiet’ does not logically, conceptually or metaphysically entail that babies are not usually quiet. There is no relation between the concepts in ‘Thora is a baby but she is quiet’ and ‘Babies are not usually quiet’, nor is it a metaphysical necessity that links these two statements. Furthermore, Target statement does not logically entail ‘Babies are not usually quiet’ because the logical roles of ‘but’ and ‘and’ are identical. Yet Target statement stands in an interesting relation to Conventional implicature whereas Descriptive meaning doesn’t. The hypothesis is that the interesting relation between the two statements is carried by the linguistic meaning of the word ‘but’. This would mean that we have a conventional implicature; as such implicatures arise in virtue of the meanings of such connectives as ‘but’ and other words.

Conventional implicatures are not cancellable and this sets them apart from conversational implicatures. You can’t say something like ‘Thora is a baby but she is quiet—yet, I do not mean to say that babies aren’t usually quiet’. The meaning of the word ‘but’ forbids you to do that. Conventional implicatures also aren’t dependent on the cooperative principle and the maxims as they are more direct, grasped more or less immediately by competent speakers of the English language.

For an example of an evaluative term, consider racial slurs. These can, when embedded in a sentence, conventionally implicate other sentences:

Target sentence:	‘Mario is a wop’
Descriptive meaning:	‘Mario is Italian’
Conventional implicature:	‘Italians are contemptible’ <sup>286</sup>

---

<sup>285</sup> Potts (2007: 666). An alternative way of presenting the conventional implicature is: ‘There is a contrast between being a baby and being quiet’. I take this difference between the two descriptions of conventional implicatures to be immaterial for what follows.

<sup>286</sup> Copp (2009b: 175, 184); Strandberg (2012: 95).

Likewise, perhaps:

Target sentence:	‘He is brave’
Descriptive meaning:	‘He doesn’t falter in the face of danger’
Conventional implicature:	‘He is morally good’

However, true as this may be, it falls short of the all-important step needed by error theorists. For at present we don’t have an explication of the non-negotiable commitment of *goodness* (or wrongness, or what have you) to the moral reality thesis. This is precisely what is not supplied by the above. After all, all we have there is a relation between a thick and a thin moral concept; not a relation between a moral concept, whether thick or thin, and the moral reality thesis. But, problematically, we can’t seem to have the following kind of conventional implicature from the thin moral concept to the moral reality thesis:

Target sentence:	‘He is morally good’
Descriptive meaning:	‘He is morally good’ (?)
Conventional implicature:	(??)

Furthermore, when sentences containing pejoratives get embedded in larger linguistic structures they typically continue to convey the speaker’s contempt for whatever it is that she focusses on: ‘If Mario is a real wop, he’ll have pasta for dinner’ continues to convey the speaker’s disrespect for Mario and Italians. This contrasts with embedding sentences containing thin moral terms into larger linguistic structures: ‘if stealing is wrong, then it’s wrong to get your little brother to steal’ does not necessarily convey that the speaker believes that stealing is wrong.<sup>287</sup> This is further reason for thinking that conventional implicatures are a poor model for understanding the error theorist’s Non-Negotiable Commitment Claim. So both kinds of implicature relations are poor models to advise the error theorist to base her account of the mode of commitment of moral discourse on (although, as I will argue in Chapter 4, both play an important role elsewhere in the defence of error theory).

---

<sup>287</sup> Strandberg (2012: 94-98).

### 3.2.3 Presupposition

The third idea is that the phrase ‘is non-negotiably committed’ is best understood as a placeholder for ‘presupposes’. This option has been tried and tested for moral error theory, albeit only rather briefly.<sup>288</sup> It is fairly customary to distinguish between two broad kinds of presupposition relations; pragmatic and semantic ones.<sup>289</sup> I discuss them, and their relation to the error theory, in the below. However, for ease of exposition, I start with a rough characterization of presuppositions that ignores this distinction.

In characterizing the notion of presupposition as neutrally as possible it is instructive to think about presupposition relations as holding between statements. Consider:

<b>Bald</b>	The present king of France is bald
<b>Not-bald</b>	The present king of France is not bald
<b>King</b>	France presently has a king

Intuitively, Bald and Not-bald contradict each other in a sense because they ascribe mutually incompatible properties about his hirsuteness to the present king of France. But they both *agree* that France presently has a king. That is, again intuitively, Bald and Not-bald both *presuppose* King.

---

<sup>288</sup> Joyce (2001: 6-9) toys with this idea but doesn’t work it out in much detail. Olson (MS: 13) writes that Sobel (MS: Chapter 13) defends a presupposition failure moral error theory. Railton (1989: 159) mentions the fact that the likely outcome of a secular account of value is “bound either to assert that our value judgments are systematically false – because they are based upon false or meaningless presuppositions – or to be revisionist in some way about the meaning of value judgments”, which, especially because he mentions this in the context of a discussion of F.P. Ramsey (an error theorist *avant la lettre*), I take to be referring to an error theory. Yet Railton goes on to investigate the revisionist option, not the false presupposition option. Shafer-Landau and Cuneo (2007) have a three page introductory piece on the moral error theory in which they only mention the presupposition approach. Lastly, although Finlay (2008a: 374; 2011: 535) *calls* the step in the error theorist’s argument where the non-negotiable commitment gets established *Presupposition*, he only engages with the arguments of Mackie and Joyce, which fall squarely in the conceptual entailment camp (Shafer-Landau (2005) is in the same predicament). Also see Timmons (1999: 78); Silverstein (2012: 1).

<sup>289</sup> Juhl and Loomis (2010: 9); Saeed (2003: 102).





So presuppositions can be triggered by various things. They can be triggered by particular words without help of a previously primed context (the Bald, Not-Bald and King case) or by the interplay of words and context (the Hammock case).<sup>293</sup> Although it is tempting to straightforwardly associate the class of *semantic* presuppositions with the former and more pragmatic ones with the latter, recent studies suggest that the relation between whether a presupposition is semantic or pragmatic on the one hand and what triggers it on the other is more complex.<sup>294</sup> Thus a presupposition can depend on context and still count as a semantic presupposition. This is good news for error theorists. Pragmatic presuppositions may be cancellable, just like conversational implicatures, because they arise via pragmatic mechanisms. They would therefore be negotiable and so a poor model for understanding Non-Negotiable Commitment Claim.<sup>295</sup>

How does this apply to the moral case? The first thing to note is that if Wrong presupposes the moral reality thesis we can't simply 'read off' this information from that statement. Contrary to Bald and Not-Bald, where King is pretty much *contained within* these statements, the information as captured by the moral reality thesis isn't likewise 'contained within' Wrong. For the moral reality thesis mentions something like 'objective prescriptivity' and that *specific* information is not 'contained within' Wrong. The information that is contained is that there exist moral properties, but that isn't specific enough. Does wrongness come with objective prescriptivity or does it not?

So in attempting to uncover the presupposition of moral statements we would have to prime an appropriate conversational context. As in other, non-moral cases, this usually happens by stipulation.<sup>296</sup> Therefore, let's prime a conversational context by stipulation.

Assume that a conversation takes place between two moralists; i.e., ordinary people debating the wrongness or otherwise of stealing. Conversations between two moralists have two very general purposes; viz., the communication of beliefs about the instantiation of moral properties and the influencing of behaviour in light of

---

<sup>293</sup> On priming contexts see Karttunen (1973: 182); Saeed (2003: 107).

<sup>294</sup> Beaver and Geurts (2011: §1).

<sup>295</sup> Stalnaker (1999); also see Williamson (2007: 88); Lycan (1999: 199).

<sup>296</sup> Strawson (1956); also see Saeed (2003: 101, 108).

those beliefs.<sup>297</sup> In this context, suppose one of them utters Wrong and his interlocutor utters Modal. Given that ordinary morality seems to so many to be invested in objective prescriptivity, what seems not to be at issue is the claim that there is such a thing as moral wrongness *understood in terms of objective prescriptivity*. What is at issue is whether moral wrongness (understood in terms of objective prescriptivity) applies to *stealing*. This means that the information that *stealing* is wrong, as entailed by Wrong (every statement entails itself) doesn't project, but that the claim *that there is such a thing as wrongness understood in terms of objective prescriptivity* does project. So we may perhaps say that the claim that there is wrongness understood in terms of objective prescriptivity is presupposed by Wrong and Modal.

There are some assumptions left unargued for in this sketch. I return to them in the next chapter in the context of critique of the presupposition option. Here I proceed by noting that although both semantically triggered presuppositions and conventional implicatures obtain in virtue of the words used in the target sentence (in the examples we used above, the target sentences were 'Thora is a baby, but she is quiet' and 'stealing is wrong'), there is a clear difference between the two. The difference is that in the case of conventional implicature, it being false that there is no opposition between being a baby and being quiet doesn't make it the case that the target sentence 'Thora is a baby, but she is quiet' is no longer apt for assessment in terms of truth and falsity. For the truth-value of 'Thora is a baby, but she is quiet' is the same as that of 'Thora is a baby, and she is quiet'. This is not true of a semantically triggered presupposition. If the presupposition is false, the target sentence is no longer capable of being true or false. Compare Bald and Not-Bald: if France doesn't have a king then both sentences are no longer apt for assessment in terms of truth and falsity.

It must also be noted that semantic presuppositions are in one respect similar to entailments and conventional implicatures: they are non-cancellable. Or at least this holds for atomic statements and their presuppositions. Consider entailments: adding 'mind you, this is not a fox' to my original statement 'this is a vixen' is not a felicitous contribution to our conversation. Likewise, denying King after having uttered Bald is not felicitous either (I can't felicitously say 'The present king of

---

<sup>297</sup> Strandberg (2012: 102-3).

France is bald, but mind you, France doesn't have a king'). However, presuppositions are cancellable when embedded.<sup>298</sup> If I say 'The present king of France is not bald—there is no present King of France!' then I *have* felicitously contributed to our conversation about the hirsuteness of the present King of France even though I've cancelled the presupposition of Not-Bald. Likewise, 'stealing is not wrong—nothing is wrong!' is a felicitous contribution to a conversation about the wrongness or otherwise of stealing even though we have a cancellation of the presupposition of Negation. Importantly though, presuppositions' limited cancellability need not be a problem for the presupposition option. Indeed, it is exactly the sort of thing that an error theorist might want to say: their limited cancellability may help to explain how various non-atomic moral statements might carry a commitment to Objective Prescriptivity without this commitment being non-negotiable in the way entailments of atomic sentences are.

As this sub-section shows, using presupposition failure as a way of understanding the error theorist's Non-Negotiable Commitment Claim yields an error theory very different from the standard error theory according to which the nature of the Non-Negotiable Commitment is defined by conceptual entailment. Furthermore, although various implicature options are quite clearly no good candidates for the error theorist's non-negotiable commitment, a second sort of entailment relation—metaphysical entailment—looks promising but has so far not even been recognized as an option for error theorists in the literature. The next chapter exploits this result to present a more nuanced picture of the prospects of both entailment options on the one hand and the presupposition option on the other. Here I continue with my list of options for error theorists; namely, her options regarding the substantive thesis.

### 3.3 The Substantive Thesis

This section discusses the various theses that moral discourse can be non-negotiably committed to. I start with the most common option, which is to argue that the moral reality thesis is the thesis that there exists a robustly understood notion of categorical moral normativity. Error theorists are free to argue that a less robust conception of normativity is morality's culprit. But this may not be a good strategy. After all, the

---

<sup>298</sup> Beaver and Geurts (2011: §3).

existence of a weak conception of morality's normativity is comparatively easy to defend. So usually error theorists go for a more robust notion of normativity. We saw Mackie's notion of prescriptivity that was meant to be 'Mackie-Objective'. We have also seen Joyce's categorical reasons of rationality. Copp has argued that there is something that makes the normativity of the moral stand apart from that of other realms, namely that each of its demands are experienced by the agent as 'welcome' or as something that is 'a rightful part of her'.<sup>299</sup> This too could feature in error theorists' arguments.<sup>300</sup>

Another option is an argument via motivational internalism or the requirement that morality would have to be able to motivate more generally. In connection with Mackie's version of error theory I argued that this option would, however, most plausibly entail an error theory of moral *beliefs* rather than morality itself.

A third option is the thesis that human beings have metaphysical free will. On this proposal, judgments about moral obligations are of central importance to moral discourse and carry a non-negotiable commitment to the claim that agents possess a metaphysically robust sort of free will. But, this version of error theory continues, there are strong arguments that this kind of metaphysical free will does not exist (cf. my discussion of Mackie's early error theory form §2.2.1).<sup>301</sup> In this thesis I set this option aside.

Fourthly, error theorists can try out different philosophical notions of morality's objectivity.<sup>302</sup> Burgess claims that moral discourse "carries [the] implication" that there exists "at least [an] intersubjectively valid and impersonal ... 'backing' for each person's moral standards".<sup>303</sup> According to Burgess, it is this 'backing' in virtue of which (true) first-order moral judgments are true. Similarly,

---

<sup>299</sup> Copp (2007: 260-3).

<sup>300</sup> Related to this is the more general option of a metaphysical framework allowing for supernatural forces, which could in turn explain the existence of categorical, robust normativity: see Hägerström (1953); Hussain (2010: 335).

<sup>301</sup> Pereboom (2001); Libet (2004); Doris (2002); Roskies (2003); Haji (1998, 2003); Newman (1981). According to Railton (2010: 305), Nietzsche, too, believed that the non-existence of metaphysical free will yields something we would nowadays call a moral error theory.

<sup>302</sup> Sayre-McCord (1986: 13-4).

<sup>303</sup> Burgess (2010: 12).

Lillehammer favours “convergence-based error theories”, where the claim in favour of this way of thinking about moral error theory is that “the issue of convergence matters in substantial moral thought.”<sup>304</sup> These are conceptions of morality’s objectivity that do not refer to mind-external entities but to the convergence of rational minds.

### 3.4 Error-riddled Moral Discourse

My task here is to explain the different commitments moral error theorists can undertake when it comes to interpreting the locution ‘flawed’ in error theory. There are two broad categories of answers: both the moral reality thesis and our moral judgments could be untrue (in various ways), and both the moral reality thesis and moral judgments could be theoretically unjustifiable (in which case we withhold judgment about their truth-value rather than positively reject them). One way of being untrue is being false, and either necessarily so (by being nonsensical or conceptually incoherent<sup>305</sup>) or contingently so. But moral judgments could also lack truth-value.<sup>306</sup> This option is most naturally combined with some versions of the Presupposition Claim.<sup>307</sup> However, Burgess endorses the ‘lacking in truth-value’-view without relying on presupposition failure:

I ... maintain that moral judgements are without truth-value ... My reason is this: It is part of the normal understanding of moral judgments that what truth-value they have is supposed to be independent of the person by whom, and the circumstances in which, they are uttered. Yet though each speaker may have quite definite (albeit unconscious) criteria for applying the term “moral,” there is not enough common to all speakers’ criteria to provide “Abortion is immoral” with a speaker-independent truth-value<sup>308</sup>

In addition to being untrue, moral judgments could also be unjustifiable.<sup>309</sup> On

---

<sup>304</sup> Lillehammer (2004: 98). Lillehammer mentions Smith (1994) as a convergence *success* theorist.

<sup>305</sup> Recall that this is Mackie’s position according to Smith (2010: 121) and others; also see Wittgenstein (1965), Hussain (2010: 338); Shafer-Landau (2003: 82).

<sup>306</sup> Joyce (2001: 28); Burgess (2010: 8); Loeb (2008).

<sup>307</sup> Strawson (1956), also see Joyce (2001: 6).

<sup>308</sup> Burgess (2010: 8).

<sup>309</sup> Kalderon (2005: 106), Kirchin (2010: 168), Hussain (2010: 337). Garner writes that “error theory ... is not necessarily a falsity theory” (1994: 139n7).

this view, the error theorist doesn't need to argue for anything as strong as the moral reality thesis or moral judgments being determinately nonsensical, false, or lacking in truth-value. All she needs is the truth of the claim that we *cannot prove* that the moral reality thesis is true. The moral reality thesis could be unjustifiable.

### 3.5 After Moral Error Theory

There are various families of stances moral error theorists can adopt in response to their theory. I list and explain them in this subsection.

First there is the abolitionist stance.<sup>310</sup> According to it we should abolish moral talk completely. One reason for doing this is that the world “would be on balance a more salubrious place if we gave up the whole business of morality.”<sup>311</sup> We should strive for ‘sensible’ solutions to for instance the abortion and euthanasia problems without getting into the muddy waters of moral arguments and the quest for a correct solution to moral problems. These quests only lead to continuing human suffering, not to quick policies that diminish suffering.

A second option is conservatism. According to it we should continue to participate in moral discourse as usual even though it carries a flawed commitment. Conservatist stances towards our flawed moral discourse subdivide into two kinds. First there is propagandist conservatism, according to which the folk should not be told the truth about error theory. The hope is that this avoids the dreadful consequences of relinquishing moral discourse—viz., predicted increases in stealing, killing, *et cetera*.<sup>312</sup> Second, there is non-propagandist conservatism.<sup>313</sup> We can have “business as usual”—we can fully believe in morality and expect to reap its benefits, such as fewer occurrences of stealing etc. compared to a state without morality—even when the folk learn about error theory.<sup>314</sup> One way to do this without sinking into inconsistency is to formulate a metaethical argument to the extent that first-

---

<sup>310</sup> Hinckfuss (1987), Garner (1994, 2010); Nietzsche (according to Joyce 2006: 107); for further arguments in favour of abolitionism, see Mackie (1980: 154).

<sup>311</sup> West (2010: 185n3).

<sup>312</sup> Discussed but not endorsed in Joyce (2001: 214; 2005). This stance is very similar to Government House Utilitarianism (Lillehammer 2004: 104).

<sup>313</sup> Mackie (1977: 16); also see West (2010: 137); Farber (1998: 161-2); Joyce (2005).

<sup>314</sup> I borrow this term from West (2010).

order and second-order moral questions are “not merely distinct, but completely independent”.<sup>315</sup> The folk can believe error theory as long as they also understand the true relation between metaethics and normative ethics.

Thirdly, we have fictionalism, which comes in two main varieties, as explained in connection with Joyce in the previous chapter. First we have content-fictionalism.<sup>316</sup> To avoid making errors, agents are not to assert and believe the proposition ‘x morally ought to  $\phi$ ’, but the *closely related* proposition ‘*in the moral fiction*, x morally ought to  $\phi$ ’. There is also attitude-fictionalism, which is what Joyce accepts. To avoid making errors, agents are not to assert and believe the proposition ‘x morally ought to  $\phi$ ’. Indeed, they are not to assert and believe *any* proposition at all. Instead agents are to *make-believe* (or adopt another attitude relatively similar to make-believe to) the proposition ‘x morally ought to  $\phi$ ’.

Fourthly, there is revolutionary expressivism. This is the view that we should start to express pro- and con-attitudes when we are engaged in moral discourse.<sup>317</sup> This has one benefit in common with fictionalism; viz., that it typically guarantees moral motivation (we are generally motivated if we are in desire-like states). It is unclear however whether recent quasi-realist efforts to save the appearances of moral discourse on an expressivist model don’t let moral truth in again through the back-door, in which case revolutionary expressivism would collapse into a form of moral realism, which is incompatible with error theory. Given also my aim to carve out the most plausible error theory within a cognitivist framework I will not take this option into consideration in this thesis.

Fifthly, there is a view that has no clear label in the literature and which I propose to call *obsoletism*. According to it, we should render the need for morality and moral discourse obsolete. This could be done by increasing “to a sufficient degree the benevolence of men or the bounty of nature.”<sup>318</sup> This option is clearly not feasible and will it set it aside in what follows.

Finally we have revolutionary cognitivism; accepted, as I argued in the

---

<sup>315</sup> Mackie (1977: 16).

<sup>316</sup> Endorsed by Lewis (1978). Discussed but not endorsed by Oddie and Demetriou (2010: 200); Juhl and Loomis (2010: 274).

<sup>317</sup> Ridge and Köhler (forthcoming).

<sup>318</sup> Hume as quoted in Lukes (1985, p. 108). Also see Lillehammer (2004: 100).

previous chapter, by Mackie. On this view, *philosophers* are to change something about the moral propositions we express in our moral judgments—not by prefixing it with an ‘in the moral fiction’-operator but by changing our conception of what it is in virtue of which something is wrong (or schwrong). The kind of change we need is one that results in moral propositions no longer containing its flawed-making properties, whilst making sure that the sentences containing them retain as much of their usefulness as possible. This can be done by turning moral propositions into *schmoral* propositions.

### **3.6 Conclusion and Preview**

This chapter has explored different options for formulating error theory. Doing this was important because, as the previous chapter showed, existing error theories with their idiosyncratic commitments are rife with difficulties. But it is possible to argue for error theory by means of different commitments, and the remainder of the thesis shows that one such alternative formulation of moral error theory is plausible. To that end I will now consider in depth the Non-Negotiable Commitment Claim (Chapters 4-5) and the Substantive Claim (Chapter 6) in light of the options listed here.



## **Chapter 4**

# **Conceptual Entailment and Presupposition**

### **4.1 Introduction**

In this chapter I argue that error theorists are ill-advised to take conceptual entailment as their interpretation of the mode of commitment of moral discourse to the poisoned thesis of morality (the moral reality thesis). I also argue that although the presupposition option can work, error theorists are best advised to explore the metaphysical entailment option instead. It is the task of the next chapter to argue that with metaphysical entailment, error theorists can get a working account of the non-negotiable commitment of moral discourse to the moral reality thesis.

I first argue that there is an objection to both the Conceptual Entailment Claim and the Metaphysical Entailment Claim, previously unrecognized in the literature, which, if successful, gets the Presupposition Claim a dialectical advantage over the entailment views. This is because the objection doesn't affect the Presupposition Claim (§4.2).<sup>319</sup> I then argue that in fact both the Conceptual Entailment Claim and the Metaphysical Entailment Claim can respond to my objection, so that the Presupposition Claim's dialectical advantage vanishes (§4.3). With the playing field level again, I argue that the Conceptual Entailment Claim is unacceptable on different grounds (§4.4). I then argue that the Presupposition Claim can work for error theorists, but that the general commitments of presupposition relations favour exploring the Metaphysical Entailment Claim (§4.5). A final section concludes and looks ahead (§4.6).

### **4.2 A New Objection to the Entailment Claim**

In this section I formulate a new objection to the view that the non-negotiable commitment of moral discourse is to be spelled out in terms of entailment without specifying what kind of entailment it is. For as will become clear shortly, that issue

---

<sup>319</sup> *Previously* unrecognized because I formulate it in Kalf (2013).

is a moot point in this context. Before I can formulate my objection we need to be more precise about the various moral statements that comprise moral discourse.

Consider:

<b>Wrong</b>	Stealing is morally wrong
<b>Permissible</b>	Stealing is morally permissible

Wrong and Permissible are atomic moral statements. Recall from §3.2 that this means that they have the logical form ‘ $x$  is  $F$ ’ and that they are ‘moral’ because the predicate  $F$  is a moral predicate. It is possible to have the following sense of permissibility in mind when saying that something is permissible: ‘stealing is *positively morally* permissible; i.e., stealing has *the property of being permitted by morality*’. But sometimes when people say that something is permissible they mean: ‘stealing, like everything else, is permissible because there is no morality’. In that case we have:

<b>Permissible*</b>	Stealing is permissible, but not <i>morally</i> permissible or permissible <i>according to morality</i> , because there is no morality
---------------------	--

Permissible\* might be a statement that we use after the truth of moral error theory, since it doesn’t attribute moral properties to things, and is therefore not a moral statement. The predicate ‘ $F$ ’ that is ‘permissible’ doesn’t read as ‘*morally* permissible’ but as ‘non-morally permissible’. In what follows I’ll have Permissible, not Permissible\* in mind.

It is possible to embed Wrong and Permissible under various linguistic operators, such as negation, modal and question operators. If we do this for Wrong we get three further, non-atomic moral statements:

<b>Negation</b>	Stealing is not morally wrong <sup>320</sup>
<b>Modal</b>	Stealing might be morally wrong
<b>Question</b>	Is stealing morally wrong? <sup>321</sup>

---

<sup>320</sup> Recall my earlier simplifying assumption that Negation is equivalent to ‘it is not the case that stealing is morally wrong’ and that its negation is used in its ordinary truth-functional rather than metalinguistic way.

<sup>321</sup> Although thus far I have understood ‘moral statements’ as linguistic utterances that express moral judgements understood as beliefs, from this point onwards, and for ease of exposition, I broaden my understanding of ‘moral statement’ to allow it to include non-atomic contributions to moral discourse, such as questions, that do not straightforwardly express beliefs.

A hitherto ignored question about all these different kinds of moral statements is: how well do the entailment and presupposition views fare with respect to *all* of them? Normally error theorists only focus on atomic moral statements and their negations; not on questions, modalized moral statements, *etc.*<sup>322</sup> But each of the five moral statements commits its utterer to the moral reality thesis to varying degrees (this will be argued presently) and so if it turns out that the entailment view or the Presupposition Claim accounts for this better than the other then that speaks in favour of adopting that view as our account of the Non-Negotiable Commitment Claim. It is this feature of moral discourse—the fact that it consists of many more kinds of moral statements than just atomic ones and their negations—that I will now use to argue against both entailment options and in favour of the Presupposition Claim.

Here is an intuitively plausible account of the relation between these five moral statements and the moral reality thesis. Wrong, given its atomic logical form, straightforwardly commits its utterer to the moral reality thesis. Negation straightforwardly does not commit its utterer to the moral reality thesis.<sup>323</sup> How about Modal, Question and Permissible? Consider moral nihilists, ordinary participants in moral discourse deeply sceptical about moral truth.<sup>324</sup> Nihilists are not willing to utter Modal and Question. Modal suggests that some things can be morally wrong: there is only a point in suggesting that things might be wrong if you've not already ruled out that things can be wrong. But nihilists believe that nothing is morally anything. Question leaves it open that some things are morally something: there is only a point in asking whether 'p' if you've not antecedently ruled out 'p'. But nihilists think that 'p' has already been ruled out, so they see no point in uttering Question. This behaviour of nihilists—their reluctance to utter

---

<sup>322</sup> Joyce (2007a).

<sup>323</sup> Further examples of sentences that contain moral terms but that are obviously not infected with the poisoned thesis that error theorists balk at include those reporting propositional attitude ascriptions ('John believes that stealing is wrong') and those containing conceptual truths ('murder is wrongful intentional killing'; Copp (2007: 116-21). I won't consider them here because they obviously can't be used as data in an argument that aims to decide which of the Conceptual Entailment Claim, the Metaphysical Entailment Claim or the Presupposition Claim is the error theorist's best bet to formulate the Non-Negotiable Commitment Claim with.

<sup>324</sup> Dreier (2006: 242). I use the locution 'nihilist' to talk about ordinary moralists and reserve 'error theorist' to talk about metaethicists subscribing to error theory.

Modal and Question—suggests that Modal and Question carry some kind of commitment to the moral reality thesis. The eventual account of the mode of commitment of moral discourse to the moral reality thesis should explain this.

Finally, consider Permissible. Nihilists will be willing to utter Permissible\* for obvious reasons. But they won't be willing to utter Permissible. This is because:

“facts about moral permissibility ... [are] not ... objectively prescriptive but rather, as we might say, objectively *permissive*”<sup>325</sup>

To understand why nihilists are unwilling to utter Permissible, recall that error theorists are impressed by morality's apparent *inescapable authority*: its ability to bind us to perform certain actions (authority) such that there is nothing we can do, not even altering our desires, to escape this authority (inescapability). It is built into this notion of inescapability that moral facts would have to be objective in the sense of being independent of human desires or attitudes more broadly. Otherwise we could change what is morally wrong by altering our attitudes and thus ‘escape’ morality's authority. But this means that the prescriptive force of moral considerations would have to be “simply there, in the nature of things.”<sup>326</sup> And this is intolerably queer—nothing can be objective and prescriptive at the same time. But if it is mysterious how things can, just by themselves, *demand* the performance of actions, then it is equally mysterious how things can out of themselves *permit* the performance of actions. The nihilists' reluctance to utter Permissible tells us that Permissible carries a commitment to:

**The moral reality thesis (permissible).**

There exists objective permissibility

Nihilists' unwillingness to utter Permissible has a further implication as well: they will also be unwilling to utter Negation, for Negation entails Permissible, by deontic logic. This means that although it remains true that Negation carries no commitment to the moral reality thesis understood as the moral reality thesis (which mentions reasons to refrain from stealing), we *do* find that Negation carries a commitment to the moral reality thesis understood as the moral reality thesis

---

<sup>325</sup> Olson (2011a: 79n.4).

<sup>326</sup> Mackie (1977: 59).

(permissible). For Negation entails Permissible and Permissible carries a commitment to the moral reality thesis (permissible).

The fact that people sometimes have Permissible\* in mind when they say that something is permissible does not threaten my claim that the error theorist's eventual account of the Non-Negotiable Commitment Claim needs to explain in virtue of what *Permissible* is committed to the moral reality thesis (permissible). And it doesn't seem plausible that people always have Permissible\* in mind when they say that something is permissible. The same holds for Modal. It too can be given an alternative interpretation according to which nihilists will be willing to utter it. This alternative interpretation is one according to which the 'might' in Modal gets an *epistemic* rather than a *metaphysical* reading. The nihilist can say: 'I acknowledge that I am a fallible epistemic agent and thus, for all I know, stealing *might* be wrong'. She will be willing to utter this because worries about epistemic access to (moral) reality don't imply a commitment to there *being* a (moral) reality. But again this would threaten my conclusion that Modal is committed to the moral reality thesis only if Modal is *never* used in its non-epistemic sense. But this too is a currently unsettled empirical claim (and one that is probably false).

So we need an explanation of the fact that each of the five moral statements carries a commitment to the moral reality thesis or the moral reality thesis (permissible), at least to the extent that it does. That is, the commitments of the non-atomic moral statements and Permissible may or may not be non-negotiable; I haven't argued this point either way. But even without this specific information we can already formulate the following challenge: neither the Conceptual Entailment Claim nor the Metaphysical Entailment Claim, in virtue of both being *entailment* relations, seems to be able to account for this. After all, it is a perfectly general feature of entailments that embedding atomic sentences carrying them under negation operators causes the embedded sentence to lose that entailment. Stalnaker writes:

A entails B if and only if B is necessitated by A but *not* by its denial<sup>327</sup>

Thus 'this is a vixen' entails 'this is a fox' but 'this is not a vixen' does not entail 'this is a fox'. Likewise, Negation doesn't entail the moral reality thesis. This is

---

<sup>327</sup> Stalnaker (1999: 54). Also see Saeed (2003: 103); Simons (2006: 357).

problematic because Negation *is* committed to the existence of a moral reality: it entails Permissible, which carries a commitment to the moral reality thesis (permissible). But it has not yet been argued that the latter commitment is an instance of an entailment. So here is an explanatory task for the entailment view: explain how Permissible *entails* the moral reality thesis (permissible) and hence how Negation entails the moral reality thesis (permissible).

In addition to Negation and Permissible, consider Modal and Question. Modal and question operators are also *entailment-cancelling operators*; e.g., ‘stealing might be morally wrong’ does not entail that it *is* wrong, and hence does not *entail* there is a categorical reason not to steal. This is also problematic, for as we saw Modal and Question do carry a commitment to the moral reality thesis. Conclusion: the entailment view is extensionally inadequate, and indeed very much so.

My reasoning here is open to two objections. First, advocates of the entailment view can “treat [moral discourse] as a term of art ... meaning a widespread linguistic practice of uttering atomic [statements]”.<sup>328</sup> If error theorists treat moral discourse in this way then my explanatory challenge no longer applies to them. For then moral discourse, by hypothesis, no longer contains non-atomic moral statements. But to make this move is to abandon the aim of providing an error theory about *ordinary* moral discourse, which consists of lots of other kinds of moral statements besides atomic ones. Yet that was the nature of the game. This objection doesn’t stick.

The second objection can be understood as an attempt to *explain away* the difficulties that Modal, Question *et cetera* present. The friend of either entailment option can

agree that people are not always the best judges of what they are talking about” and insist that the telling evidence is how nihilists respond to cases where the commitment to [the moral reality thesis] is explicitly denied, not how they respond to [complicated statements like Modal]<sup>329</sup>

In reply, I agree that the folk may be mistaken about their own discourse, thinking that a certain non-atomic statement like Modal carries a commitment to the moral reality thesis even though, by hypothesis, it doesn’t. But I insist that in that case we

---

<sup>328</sup> Joyce (2007a: §4); also see Sinnott-Armstrong (2006: 34-6).

<sup>329</sup> Joyce (2011a: 527).

need an explanation of why the folk are mistaken in the *systematic* way that they are. We have a large number of statements that *seem* to carry a commitment to the moral reality thesis —Modal, Question, Permissible, Negation via Permissible, and probably many more—and if none of them in fact carries that commitment then the fact that they *appear to* requires explanation. Either way, advocates of the entailment view have explanatory work to do. This objection doesn't stick either.

To sum up what we have so far, the entailment view is unable to account for the fact that various moral statements embedded under entailment-cancelling operators seem to carry commitments to the moral reality thesis. It also has yet to explain how, exactly, Permissible (and hence Negation as it logically entails Permissible) is committed to the moral reality thesis (permissible). I will now argue that the error theorist's back-up plan, the Presupposition Claim, does a much better explanatory job here.

Recall that for presuppositions between moral statements and a statement specifying the moral reality thesis, a suitable conversational context needs to be primed. This is because if Wrong, Negation, Modal, Question and Permissible presuppose the moral reality thesis or the moral reality thesis (permissible), we may not be able to (or perhaps can't) simply 'read off' this information from these statements. For the moral reality thesis mentions objective prescriptivity and the moral reality thesis (permissible) mentions objective permissibility and that *specific* information is not contained within our five moral statements.

So let's prime an appropriate conversational context by stipulating that our conversation takes place between two moralists; ordinary people debating the wrongness or otherwise of stealing. One of them utters Wrong and his interlocutor utters Modal. On this assumption, given that ordinary morality seems to so many to be invested in objective prescriptivity, what is not at issue is the claim that there is such a thing as moral wrongness *understood in terms of objective prescriptivity*. What is at issue instead is whether moral wrongness understood in terms of objective prescriptivity applies to *stealing*. This means that the information that *stealing* is wrong, as entailed by Wrong (every statement entails itself) doesn't project, but that the claim *that there is such a thing as wrongness understood in terms of objective prescriptivity* does project. This is exactly what we intuitively find (or at least error theorists and non-naturalist realists argue) and so we may perhaps say that the claim that there is wrongness understood in terms of objective

prescriptivity is presupposed by Wrong and Modal. Applying this to Question, we find that, at least when the intonation is on stealing (as in ‘is *stealing* wrong?’) the information that there is wrongness understood in terms of objective prescriptivity projects.<sup>330</sup> For that information is not at issue if you wonder about the wrongness of *stealing*. And something similar holds for Permissible. If I utter Wrong and you utter Permissible then *that* there are things that have moral properties is not at issue and hence this information projects; for the only thing that is at issue is whether *stealing* has the property of being permissible or of being wrong.

However, in fact whether the information that projects must be understood in connection with categorical reasons, objective permissibility, or a mixture of these is not clear. That is, it seems that saying that the information that projects mentions not objective permissibility but categorical reasons instead would work equally well. Or indeed we can imagine less robust accounts of morality’s normativity or even claim claims about free will or rational convergence (cf. §3.3) to accompany the projected information ‘that there exist moral properties’. So although the Presupposition Claim seems to work *in principle* when what we have in mind is extensional adequacy—viz., an explanation of why a whole raft of different kinds of moral statements seems to be committed to the moral reality thesis to various degrees—we also need to tighten our explanation. For, again, what we have at present is a view on which the mode of commitment of moral discourse to the moral reality thesis leaves it open what the precise *content* of the moral reality thesis is: objective permissibility, categorical reasons, or what have you. I assume that an eventual, more fully developed account of the Presupposition Claim can work this out. The important point though is that it doesn’t seem *impossible* for the friend of the presupposition claim to argue that Modal *et cetera* presuppose the moral reality thesis. This contrasts with the Conceptual Entailment Claim and the Metaphysical Entailment Claim. For the Conceptual Entailment Claim and the Metaphysical Entailment Claim, given that various sentential operators such as modal operators provide entailment-*cancelling* embeddings for the atomic statements they are based upon, it is impossible to explain why statements like Modal *et cetera* are committed to the moral reality thesis. So, to conclude, the Presupposition Claim has an important dialectical advantage over the Conceptual Entailment Claim and the

---

<sup>330</sup> When the intonation is on wrong then Question could be heard as asking the question whether anything is wrong at all.



Metaphysical Entailment Claim. The Presupposition Claim can whereas the Conceptual Entailment Claim and the Metaphysical Entailment Claim cannot provide, at least in principle, an extensionally adequate account of the mode of commitment of the various moral statements that comprise moral discourse to the moral reality thesis.

### 4.3 Entailment and Pragmatics

Or does the Presupposition Claim have this advantage? I will now argue that in fact it doesn't. Our discussion of the entailment view in the previous section shows that it consists of the following theses:

<b>Positive Thesis</b>	Atomic moral statements (except those about permissibility) entail the moral reality thesis (Permissible entails the moral reality thesis (permissible))
<b>Negative Thesis</b>	Non-atomic statements do not entail the moral reality thesis

Negative Thesis bears its name because it says that the entailment view is only committed to the, precisely, negative thesis that non-atomic statements do not *entail* the moral reality thesis. But this leaves it open that another relation apart from entailment can be used to close the gap.

Here is a proposal. Modal, Question and Permissible keep the existence of a moral reality open as a live option because they *conversationally implicate*:

<b>Reasons</b>	There exists objective permissibility as well as objective prescriptivity and hence, in virtue of the latter fact, categorical moral reasons <sup>331</sup>
----------------	---

Recall from the previous chapter that conversational implicatures are standardly understood along Gricean lines. Grice explains implicatures, or how statements convey information without that information being part of their meaning, with the aid of two phenomena: the cooperative principle and conversational maxims. The cooperative principle says:

[m]ake your conversational contribution such as is required at the stage at which it occurs, by the accepted purposes or direction of the talk exchange in which you are engaged<sup>332</sup>

---

<sup>331</sup> For a similar proposal see (Olson 2011a: 69-70).

The maxims tell you what to do to meet the principle; e.g., the maxim of *quantity* tells you to make your contribution neither less nor more informative than is required, and the maxim of *relation* tells you to make your contribution relevant. According to Grice, we are justified in assuming that people conform to the cooperative principle and hence to the maxims (absent special circumstances such as when people are acting).

In our case, the implicature between Modal, Question and Permissible on the one hand and Reasons on the other holds because, looking at Permissible for example, we have in our society a moral standard according to which morally relevant actions are either permissible or forbidden (but not both) and because to state that information explicitly would be to violate the Gricean maxim of quantity, which forbids us to be overly informative when we speak. Here is the calculation of the implicature. You utter Permissible. I am justified in thinking that you abide by the maxim of quantity as your means of abiding by the cooperative principle, and since I know that it is part of our conversational context that we have a moral standard in place according to which various morally relevant actions are either morally prescriptive or morally permissive, I know that your utterance of Permissible carries the implicature that there are *other actions* that carry categorical moral reason-giving force.

This explains why nihilists are not willing to utter Permissible. By uttering Permissible they commit themselves, via its *implicature* to Reasons, to the existence of categorical reasons. But nihilists can also reveal their hands and clarify that they mean to say that stealing isn't wrong because nothing is morally anything. Although Wrong still entails the moral reality thesis in that case, the implicature of Permissible to Reasons gets cancelled, and so nihilists will be happy to say that something is permissible. For then it has become clear that what nihilists intend to convey is Permissible\*. Note that Negation, since it (logically) entails Permissible, gets a similar treatment. Nihilists are initially unwilling to utter it because it expresses allegiance to the moral game: Negation entails Permissible which conversationally implicates Reasons. But when that conversational implicature gets cancelled, nihilists are willing to utter Negation. They might say: 'yes, stealing is not wrong, and hence (non-morally) permissible, because nothing is morally anything'.

---

332 Grice (1989: 6).

Something similar again holds for Modal and Question; these are also statements that nihilists are initially unwilling to utter as they are suggestive of things having moral status, and thereby trigger recognition of Reasons. And nihilists will be happy to utter Modal and Question once they have cancelled these implicatures. For then Modal may take up its epistemic reading and Question could be used as a means of testing the allegiance of others to either moral realism or moral skepticism. ‘Is stealing wrong?’ will elicit a positive response from a moralist and the following, implicature cancelling response from the nihilist: ‘No, stealing is not wrong—nothing is wrong!’

By incorporating implicatures into their theory, friends of either the Conceptual Entailment Claim or the Metaphysical Entailment Claim have the resources to explain the commitments of the various kinds of moral statements to the moral reality thesis, just like the Presupposition Claim can. So the playing field is level again, but, importantly, we are not back to square one. We have gained important insights into the kinds of moral statements that carry a commitment to the moral reality thesis, and how, exactly, entailment and presupposition options explain this. But which one shall we choose? Entailment or presupposition? I will start by arguing that the *conceptual* entailment option is not the error theorist’s best bet.

#### **4.4 Against Conceptual Entailment**

In this section I argue that error theorists shouldn’t model their account of the mode of commitment of moral discourse to the moral reality thesis on conceptual entailment, given that metaphysical entailment is also available. The reason for this is that, compared to the metaphysical entailment claim, a conceptual entailment claim is relatively difficult to defend, and for two reasons, as I will now explain.

In §3.2.2.3 I explained that the main difference between both conceptual and metaphysical entailment relations concerns what it is in virtue of which the entailment holds (concepts vs. *a posteriori* considerations). This difference yields a concomitant difference in our epistemic access to both entailment relations. Conceptual entailments can be discovered by looking for connections between concepts whereas for metaphysical entailment *a posteriori* investigation is required.

We can add to this a further difference. Every conceptual entailment is also a metaphysical entailment, but not *vice versa*. If it is part of the applicability

conditions of the concepts VIXEN and FEMALE FOX that they both refer to the same object (female foxes) then the propositions ‘this is a vixen’ and ‘this is a female fox’ not only conceptually but also metaphysically entail each other. This is because the re-written propositions ‘this is a female fox’ and ‘this is a female fox’ obviously metaphysically entail each other. However, the relation of metaphysical entailment that obtains between ‘this is water’ and ‘this is H<sub>2</sub>O’ does not bring forth a relation of conceptual entailment between those two propositions. For it is not part of the *meaning* of ‘this is water’ that that thing is H<sub>2</sub>O.

This suggests that conceptual entailment relations relevant to moral error theory are more difficult to defend, and for two reasons. First, since not all metaphysical entailments are conceptual entailments and since all conceptual entailments are metaphysical entailments as well, aiming for a metaphysical entailment simply enlarges the error theorist’s chance for success.

A second reason why conceptual entailments are more difficult to defend than metaphysical entailments has to do with the first difference, which concerned our epistemic access to both entailment relations.

In the context of a discussion about the Conceptual Entailment Claim, Finlay asks:

[t]he difficult question ... how is the [... conceptual] content of our language and thought determined? [One idea] turns on considerations of people’s reflective understanding of their moral thought and speech, and of what they may be conscious of when they engage in this thought and speech. This evinces an assumption of the truth of a local form of *content-internalism*: what we mean morally is fixed by something internal to our mental states, particularly our intentions ... While moral content-internalism doesn’t entail that our reflective understanding of our practices is infallible (we can be mistaken about our own intentions and can thereby misunderstand our own concepts and language), it is what supports the view that reflective evidence can be decisive in establishing which theory of moral judgment is correct<sup>333</sup>

Error theorists tend to like content-internalism.<sup>334</sup> Recall Joyce’s *translation test-argument*. Take a contender moral concept that doesn’t carry a commitment to some kind of inescapable authority and see if (competent) users of moral language would

---

<sup>333</sup> Finlay (2008a: 362). Also see Nichols (2004: 192-3).

<sup>334</sup> Also see Shafer-Landau (2003: 67).

be happy to accept that concept as a *moral* rather than a *schmoral* concept. If they are happy to accept it as a moral concept then a concept like MORALLY WRONG doesn't entail INESCAPABLE AUTHORITY and the contender moral concept is indeed a moral concept. If however they aren't happy to accept this then MORALLY WRONG does entail INESCAPABLE AUTHORITY. The translation test argument relates to content internalism in the following way: if *we* think that INESCAPABLE AUTHORITY is part of the concept MORALLY WRONG because *we* decide and intend to use MORALLY WRONG only when there is inescapable authority afoot, then (ignoring for ease of exposition of the relevant point I want to make here the complexities introduced by our fallibility) INESCAPABLE AUTHORITY *is* part of MORALLY WRONG.

Here then is a problem for the Conceptual Entailment Claim. It ignores content-externalist accounts of the conceptual content of our language and thought, according to which the meaning and reference of some of the words we use are not solely determined by the ideas we (collectively or individually) associate with them.<sup>335</sup> But this is a gratuitous assumption in the context of arguing for error theory. For if content-externalism is true then it might be that INESCAPABLE AUTHORITY can be safely removed from MORALLY WRONG even if the folk are reluctant to accept that. That is, if content-externalism is true then moral concepts may well be truly applicable to aspects of the world that do not exhibit categorical normativity, and this may be so even when the folk think that this shouldn't be possible.

This gets us an Alternative Explanation Objection to the Conceptual Entailment Claim: *even if* moral discourse evinces a clear commitment to categorical reasons, it still doesn't follow that this commitment "contaminates" moral concepts.<sup>336</sup> For there is an alternative explanation of the fact that moral discourse evinces a commitment to categorical reasons. This is that the folk have *false beliefs* about the referents of the moral terms which stand for moral concepts. The folk *think* that these referents have to have categorical reason-giving force. So (i) they don't accept contender moral concepts that do not have categorical reason-giving

---

<sup>335</sup> Lau and Deutsch (2002: §2).

<sup>336</sup> The word 'contaminates' comes from Finlay (2008a: 347). Also see Field (1973); Brink (1989: Ch. 6); Gampel (1996); Copp (2007: 203-8, 232-6); Cuneo (2012: 123n25). Sainsbury's (1998: 141-2) and Baillie's (2000: 15) interpretations of Hume's ethics follows much this pattern too.

force built into them as *moral* concepts and (ii) they feel enfeebled in their criticism of other people's behaviour when they can't use concepts that have categorical reason-giving force built into them, *et cetera* (here I am referring again to parts of my discussion of Joyce's attempts to settle the Non-Negotiable Commitment Claim from §2.3.1). But according to the alternative explanation, moral terms are free from error and successfully refer to moral properties that do not have categorical reason-giving force. The folk just have false beliefs about, or have wrong *referential intentions* for, moral terms that stand for moral concepts.

How could this happen? The explanation that opponents of error theory tend to give is that the referents of terms, including moral terms, can be opaque to their users despite their users' competence with them.<sup>337</sup> For:

It is not ... plausible to claim that any competent user of the predicate 'moves' knows that it expresses the concept of a relation rather than the concept of a monadic property ... Some perfectly competent possessors of the concept of motion were unaware that the only truths there are about motion are relational ones<sup>338</sup>

After all, had it not been possible for competent users of terms to fail to know what concept it expresses, exactly, and thus had it not been possible for them to fail to know what the constitutive elements of those concepts are, then entire discourses would have to be relinquished to systematic falsehood. Error theories of such discourses would become our only choice. This holds for the terms 'moves', 'water' and 'mass'. Harman writes:

Before Einstein, judgments about mass were not intended as relative judgments. But it would be mean-spirited to invoke an "error theory" and conclude that these pre-Einsteinian judgments were all false! Better to suppose that such a judgment was true to the extent that an object has the relevant mass in relation to a spatio-temporal framework that was conspicuous to the person making the judgment, for example, a framework in which that person was at rest<sup>339</sup>

Which way do moral terms and concepts go? Can they be said to be free of error like 'mass', 'water' and 'moves' or do they contain errors like 'phlogiston' and 'witch', for which error theories are obviously appropriate? The objection that I am developing is that error theorists haven't done enough to argue that moral concepts

---

<sup>337</sup> Jackson (2012: 79-82).

<sup>338</sup> Boghossian (2006: 15).

<sup>339</sup> Harman and Thompson (1994: 4). Also see Horgan and Timmons (2002: 74, 79-80); Putnam (1988: 9-10).

are like the concepts WITCH and PHLOGISTON. True, we have seen Mackie's and especially Joyce's attempts to argue that moral concepts go that way. Mackie has tried to conclude from our patterns of reactions to a perceived need to abandon belief in moral values that objective prescriptivity is part of moral concepts. Joyce employed Mackie's Platitude, the translation test argument and various versions of the use argument. But we have already seen that these attempts have failed (cf. Chapter 2). Moreover, they assume the truth of content-internalism without making this explicit and without defending this assumption. This is a gratuitous assumption in the context of arguing for error theory. And, furthermore, content-internalism is difficult to defend. For recall that even moral error theorists have to assume that for many terms, including 'moves', 'mass' and 'water', a successful content-*externalist* story is available. Otherwise we end up with the undesirable consequence that we also have to be error theorists about motion-, mass- and water-discourse. But what justifies accepting content-internalism other than that it can give error theorists what they want? Indeed, one might think that since accepting content-internalism about moral discourse leads to an error theory of that discourse, we have reason to *reject* it.

Importantly, the Metaphysical Entailment Claim survives this critique because it doesn't apply to it. After all, according to the Metaphysical Entailment Claim, 'stealing is wrong' metaphysically entails 'there is robust categorical moral normativity' even though that information need not be built into the concept MORAL WRONGNESS. This means that what we mean morally can be fixed by something *external* to our referential intentions. So, with the Metaphysical Entailment Claim, error theory is no longer open to this particular objection because the advocate of the Metaphysical Entailment Claim *agrees* with that objection that there are aspects of the world that obtain external to agents' minds that can affect which statements stand in entailment relations with other statements. It is just that this kind of error theorist who accepts the Metaphysical Entailment Claim argues that when we move away from the thought that "what we mean morally is fixed by something internal to our mental states, particularly our intentions," and broaden our horizon, we find that moral discourse can't be *moral* discourse without a robust kind of categorical moral normativity.

Friends of the Conceptual Entailment Claim could respond to my objection by saying, with Joyce, that although:

there are empirical disclosures to be made about water [for example that water is H<sub>2</sub>O], it could not be discovered that “water” is a mass term for the time between lunch and dinner. It is our intentions (considered collectively) to use the term in a certain way that preclude this possibility<sup>340</sup>

Joyce’s idea is that any entailment relation needs *some* (collective) referential intentions to ensure extensional adequacy of the relation on offer. But the need for some referential intentions is consistent with content-externalism (just like an externalist causal theory or reference is consistent with an initial baptizing of a term).<sup>341</sup> Moreover, since, as Joyce himself says, the referential intentions may be *collective* rather than *individual* referential intentions, it is not even true that the necessity of incorporating referential intentions in figuring out which entailment relations there are requires us to move away from content-externalism. Finally, there is also no guarantee that the (collective) referential intentions which constrain reference assignment for moral terms will have to carry a commitment to categorical reasons (or anything else that is useful for a moral error theory for that matter).

To sum up, I have discussed two considerations that put the Conceptual Entailment Claim in bad light. First, if error theorists use the Conceptual Entailment Claim rather than the Metaphysical Entailment Claim then they needlessly limit their chances for success, as every conceptual entailment is also a metaphysical entailment but not *vice versa*. Second, using the Conceptual Entailment Claim requires settling the debate about content-internalism about what fixes what we mean morally, or if that is not possible then it requires *assuming* content-internalism. But both of these options are bad news for error theory. The first option is hard to defend, given that we know that many terms in our natural languages can successfully refer even if the folk have false beliefs about the constituent parts of the concepts that these terms express. The second option is intellectually dishonest. With the Metaphysical Entailment Claim however error theorists can by-pass these issues as the Metaphysical Entailment Claim is consistent with content-externalism about what fixes what we mean morally. Better then to relax the requirement on entailment relations that they involve connections between concepts to something more akin to the Metaphysical Entailment Claim. Indeed, I will argue in the next chapter that the Metaphysical Entailment Claim can give error theorists a working

---

<sup>340</sup> Joyce (2001: 97). Also see Parfit (2012 Vol2: 301-2).

<sup>341</sup> Kripke (1972).



argument for their Non-Negotiable Commitment Claim. But first let me argue that the Presupposition Claim is not a good option for error theorists either.

## 4.5 Against Presupposition

In §4.2 and §3.2.3 we saw that the Presupposition Claim makes at least three assumptions that stand in need of defence. One was that we can specify a conversational context in which the presupposition that there is a moral reality (understood in a sufficiently robust and explicit way congenial to error theory) arises even when that context is described *sufficiently minimally* in order to avoid the charge that we are building into the context what we want to get out of it. This is not evidently possible. Recall that as opposed to the case of Bald, Not-Bald and King, where the presupposition that there is a king (King) can be read off both Bald and Not-Bald pretty straightforwardly, the presuppositions of moral statements, if any, cannot likewise be straightforwardly read off them. In line with standard solutions to this problem I then suggested that what's needed is the priming of an appropriate conversational context: we need to get more information about the context of the conversation to be able to say what, if any, the presuppositions of the statements involved are. I primed one conversational context—one in which we have two moralists debating the wrongness or otherwise of stealing—and asked what the presuppositions of these statements would be. I suggested that one bit of information that may not be at issue in that conversation (recall the projection test for presuppositions according to which a piece of information projects so long as it is 'not at issue' in that conversation, and according to which what is thus projected may count as a presupposition of that discourse) *may be* that there is categorical moral normativity. The problem is: lots of other bits of information are consistent with this story as well. Take: there is *hypothetical* moral normativity. That information too can be said to be not at issue when one person utters Wrong and the other utters Negation in response. Which of these is the presupposition of Wrong and Negation? Because we primed the conversational context sufficiently minimally to avoid the objection that we are building into the context what we want to get out of it, this is not clear.

So let's put more information into the context. Say, for instance, that both interlocutors are very religious and believe that moral rules are God-given, set in stone, strongly normative, that they entail categorical reasons, *et cetera*. With *that*

information specified as the conversational context the suggestion that the information that projects is hypothetical moral normativity, not categorical moral normativity, seems out. However, our victory is very much a Pyrrhic victory: we only got the projected information that there must be *categorical* moral normativity out of the exercise, or so it very much seems, because *we* put that information into the conversational context to begin with.

This dilemma between beefing up the conversational context in order to get a clear presupposition but risking assuming what we set out to prove *versus* describing the conversational context sufficiently minimally to avoid that risk but then landing ourselves with an indeterminate presupposition is a big problem for the presupposition approach. Note too that the friend of the Metaphysical Entailment Claim doesn't need to prime conversational contexts in this way. She is therefore better off than the friend of the Presupposition Claim.

The second assumption, leading on from the first, was that in a conversation in which Wrong and Permissible get uttered, it is not clear which of the moral reality thesis or the moral reality thesis (permissible) gets presupposed, and this too would need to be settled. Again, this problem has everything to do with the problem of priming an appropriate conversational context, and again, this task is one that a friend of the Metaphysical Entailment Claim doesn't have to perform.

The friend of the presupposition approach can respond to these problems as follows. She can say that I've exaggerated the need to prime a conversational context. I talked about priming a context for talk exchanges between individual agents. But it is also possible to look at the shape and function of moral discourse as a whole, and to take the natural prediction that what is taken for granted in normal contexts of sincere moral discourse to be moral discourse's presupposition. This allows the friend of presupposition to avoid priming conversational contexts altogether. All we need, on this picture, for teasing out the 'poisoned presupposition' of moral discourse are general reflections on platitudes about moral discourse, thought-experiments, intuitions and the like.<sup>342</sup> The benefit of this suggestion is that we can get a presupposition relation favourable to error theory on the table without priming any conversational contexts, and thus without running head-on into the trouble of reading into moral conversations what we want to get out of them.

---

<sup>342</sup> See Finlay (2011: 535) on the terminology 'poisoned presupposition'.

However, still, the friend of metaphysical entailment has the upper hand. For notice that this way of looking at and arguing for ‘presupposition’ relations is virtually identical to looking at and arguing for ‘metaphysical entailment’ relations. After all, in arguing for metaphysical entailment relations, I suggested in §3.2.2.3, error theorists should also invoke ‘general reflections on platitudes about moral discourse, thought-experiments, intuitions and the like’. But then talk of a *presupposition* rather than an *entailment* relation is potentially distracting. For it looks as though it is much easier to make sense of a relation of entailment between an atomic moral proposition and the proposition about objective prescriptivity than it is to make sense of a relation of presupposition between these two propositions. The reader might recall from §3.2.3 the complexities involved in presupposition relations, including the issue of semantic versus pragmatic presuppositions and our epistemic access to them, to name just two. Error theorists hoping to settle their Non-Negotiable Commitment to something like objective prescriptivity by means of mustering *a posteriori* reflections on the nature of moral wrongness would be well-advised to avoid these complexities by interpreting the relation that they find between a moral proposition and a proposition about objective prescriptivity in terms of entailment. This is not a knock-down argument against the presupposition option. To the contrary; it might still be the case that for all this thesis shows, in fact researching the presupposition option gives error theorists the best account of moral discourse’s Non-Negotiable Commitment. But I do think that given the aforementioned complexities with presupposition and the relative straightforwardness of entailment, error theorists should try to work on metaphysical entailment first.

The final assumption that we had to make to make the Presupposition Claim work was that the Presupposition Claim would have to be able to explain in virtue of what the presuppositions of Modal and Question are *cancellable with felicity*, as it looks like they are. For remember that it makes good sense to say ‘no, it is not the case that stealing might be wrong—nothing is wrong!’ Here however the advocate of the Metaphysical Entailment Claim faces a similar problem. She has to explain how the *implicatures* of Modal and Question are *cancellable with felicity*. The advantage of metaphysical entailment over presupposition here, however, is that again its pragmatic account of cancellability with felicity seems to be better understood than that of the various theories of presupposition. Again though this is

not a knock-down argument against presupposition. For all this thesis shows friends of the Presupposition Claim can tell a perfectly convincing story about this. But it does indicate that error theorists should seriously consider looking at metaphysical entailment before considering presupposition.

## 4.6 Conclusion and Preview

In this chapter I have argued that there is a currently unaddressed objection to using entailment relations between moral statements and the moral reality thesis to make sense of the error theorist's Non-Negotiable Commitment Claim. The objection is that although many of the kinds of moral sentences that comprise moral discourse are correctly, from an intuitive point of view, either debunked or vindicated by an entailment claim, there is also a class of moral sentences that the entailment options do not debunk even though these sentences intuitively ought to be debunked. I argued that there is reason to believe that this objection favours the Presupposition Claim over either the Conceptual Entailment Claim or the Metaphysical Entailment Claim. I also argued that this objection in fact fails because the friend of the entailment claim can use pragmatic mechanisms to account for the relevant data. But this didn't get us back to square one. Engaging with the hitherto unrecognized objection showed us which kinds of moral statements (atomic statements, non-atomic modalized statements, non-atomic negated statements, etc.) probably carry the commitment to the moral reality thesis and to what degree. This then allowed me to object to the Presupposition Claim on the grounds that the assumptions it must make to account for these data are at least *prima facie* more difficult to defend than the assumptions of the friend of the Metaphysical Entailment Claim. I concluded from this that it would make sense for the error theorist to study the Metaphysical Entailment Claim rather than the Presupposition Claim. I also argued that there are difficulties internal to the practice of proposing relations of *conceptual* entailment that the friend of the Metaphysical Entailment Claim doesn't have to deal with. This gives us *prima facie* reason to study the Metaphysical Entailment Claim rather than the Conceptual Entailment Claim. The next chapter does precisely that, and argues that using the Metaphysical Entailment Claim gets us a working argument for the claim that ordinary moral discourse carries a non-negotiable commitment; namely, to a robust kind of categorical moral normativity.

# Chapter 5

## Metaphysical Entailment

### 5.1 Introduction

His own efforts to argue for moral error theory notwithstanding, Joyce believes that

no error theorist has yet succeeded in making [the case for the Non-Negotiable Commitment Claim] persuasively<sup>343</sup>

In previous chapters I argued that thus far error theorists have tried to establish their Non-Negotiable Commitment Claim either by arguing for the Conceptual Entailment Claim or by arguing for the Presupposition Claim. I argued that the Conceptual Entailment Claim is rife with difficulties and that the Presupposition Claim is heavily underexplored and, compared to the Metaphysical Entailment Claim, probably unnecessarily complicated. This explains and corroborates Joyce's remark. In this chapter I aim to settle the Non-Negotiable Commitment Claim via the Metaphysical Entailment Claim. the Metaphysical Entailment Claim allows error theorists to argue that objective prescriptivity (or what have you) is part of the truth-conditions of moral discourse without making this analytic of moral discourse.

As far as I know the Metaphysical Entailment Claim has so far not been explored at all. Lillehammer comes close when he writes that:

the [non-negotiable commitment] claim is fundamental to some moral thought (although I don't claim that [the] commitment ... is ... [an] analytic ... condition for basic competence with moral terms)<sup>344</sup>

This looks like the kind of metaphysical entailment that I'm after: a commitment to something like objective prescriptivity is fundamental or essential to moral discourse although it is not analytic of moral discourse (or our competence with it). However, having made this claim, Lillehammer doesn't explore this option in much detail.

---

<sup>343</sup> Joyce (MS: 10); also see Cuneo (2012: 114-5).

<sup>344</sup> Lillehammer (2004: 99).

Instead he makes this comment to motivate error theory to a sufficient degree to justify a discussion of what we should do with moral discourse *after* error theory.

How might one go about establishing a metaphysical entailment claim? One direct strategy is to look at various platitudes about the nature and status of morality as we get them from a carefully executed *internal accommodation project* of moral discourse. With those platitudes in place we can then formulate a hypothesis about the nature and status of morality that provides the best possible fit with those platitudes. There are complexities involved that I will discuss in due course, but the guiding idea is that this hypothesis embodies the non-negotiable commitment of moral discourse. After all, the hypothesis is what undergirds and explains platitudinously true claims about moral discourse in the best possible way.

Let's be clear about the relationship between (1) what is analytic of moral discourse, (2) what is platitudinously true of moral discourse and (3) what best explains the platitudes. What I *deny* is that the non-negotiable commitment of moral discourse is analytic of it. That is, I deny that it is part of moral concepts that they can only be successfully applied to (say) objectively prescriptive moral facts and properties (Mackie), categorical reasons (Joyce), inter-subjective agreement (Burgess, Streumer, Schiffer), or what have you. What I *accept* is that the platitudes themselves can be analytic of moral discourse. For this is consistent with (3) what best explains these platitudes is some thesis about morality that we discover via *a posteriori* means rather than conceptual analysis.

I start by enumerating platitudes about the nature and status of moral discourse (§5.2). I then argue that the hypothesis that there exists categorical *normativity understood in a particular way to be explored* provides the best fit with these platitudes, and that a Metaphysical Entailment Claim conducive to error theory is therefore in the offing (§5.3). The chapter further discusses one important objection to my argument, but rejects it (§5.4). It closes with a conclusion and preview (§5.5).

## 5.2 Platitudes

In this section I discuss platitudes about the status and content of morality. The claims I am about to present count as platitudes on account of the fact that they are very much ingrained in folk moral thought. Recall from §1.1 my method of *quasi-*

empirical research where metaethicists make claims about what is platitudinously true about moral discourse on the basis of their own intuitions and experiences. Although it is always possible to overthrow platitudes on theoretical grounds, doing so comes with the cost of accepting a theory that cannot handle certain platitudinously true claims. I will rely on fairly minimal platitudes—platitudes that are formulated in such a way that they allow for exceptions and are consistent with various different theoretically more loaded explanations. The argument of this chapter is that one theoretical interpretation provides the best fit with these platitudes (and the argument of the next chapter is that this commitment isn't made true by the world).

One further comment applies. The platitudes that I formulate involve moral oughts or the deontic aspects of moral discourse only. Whether my argument generalizes to evaluative aspects of moral discourse (and whether it is a big problem if it doesn't) will be discussed in §7.4.

Here, then, is the first platitude:

#### **Desire-Independence**

If an agent morally ought to  $\phi$ , then that agent morally ought to  $\phi$  regardless of whether she feels like  $\phi$ -ing (at the time) and regardless of whether  $\phi$ -ing would further any of her goals even if she thought about  $\phi$ -ing for a long time<sup>345</sup>

The idea behind Desire-Independence is that whether it is true of agents that they morally ought to  $\phi$  neither depends on whether they want to  $\phi$  at the time of assessment nor on whether  $\phi$ -ing furthers any of the agent's goals. Various infamous characters in history and the history of philosophy morally ought not to have done what they did, and the folk believe that this is true regardless of what these rogues wanted or what their goals were. These thugs include, but are not limited to, Genghis Kahn, Caligula, Nero, Hume's sensible knave, Gyges, and Thrasymachus.<sup>346</sup>

This initial characterization requires various qualifications. First, Desire-Independence only concerns moral oughts. That is, it only says that if Jack morally ought to refrain from killing then he morally ought to do that regardless of what he

---

<sup>345</sup> "Common opinion agrees with Kant in insisting that a moral man must accept a rule of duty whatever his interests or desires" Foot (1972: 306).

<sup>346</sup> The reference to Hume is to his (1739-40: 3.2.1.9). Gyges and Thrasymachus feature in Plato's *Republic*.

feels like doing at the time and regardless of whether he still desires to kill after careful deliberation. Desire-Independence leaves it completely open whether or not *ignoring this moral ought entails making a rational mistake*. Killing could be wrong *relative to* the rules of morality whether or not one feels like killing. But all the same it could be a further question whether it is always *irrational* to kill.<sup>347</sup> If one thinks that being practically irrational consists in ignoring a *reason* that one has, then the question is whether the categorical wrongness of stealing translates into all agents also having a *reason* not to steal. (But one could also hold that being practically irrational consists in ignoring rational requirements instead, in which case the question becomes whether the categorical wrongness of stealing translates into all agents being under rational requirements not to steal).

The second qualification about Desire-Independence is that if we think that moral oughts are not rational oughts then we must make further distinctions. For if it is not true that moral oughts are always rational oughts then we must ask: how should we then understand the way in which morality is normative? For this kind of moral ought is, by hypothesis, not normative because it is a rational ought. But then then *in what sense* or *in what way* is it normative?<sup>348</sup>

Rational oughts are in good sense straightforwardly normative. If I'm *irrational* when I'm doing something morally wrong then I'm not living up to the standards of rationality. But what are the other options? Restricting ourselves to moral normativity we can note the following possibilities:

**Rule-Involving Oughts**

Moral oughts are normative in virtue of being rules, and rules are normative in virtue of distinguishing between what is *correct* and *incorrect* or what is *allowed* and *disallowed*<sup>349</sup>

**Motivational Oughts**

Moral oughts are normative in virtue of entailing actual or possible motivation, such that when a putative ought, known by an agent, doesn't actually or possibly motivate, it fails to be an ought<sup>350</sup>

**Attitudinal Oughts**

Moral oughts are normative in virtue of involving attitudes of approval or disapproval to our own or other people's actions. Attitudes of approval or

---

<sup>347</sup> Foot (1972); Railton (1986); Brink (1984, 1989).

<sup>348</sup> Copp (2012: 25).

<sup>349</sup> Parfit (2012 Vol1: 144).

<sup>350</sup> Parfit (2012 Vol2: 268); Copp (2007: 258-60).



disapproval are often motivationally efficacious, but this feature isn't what guarantees the relevant oughts' normativity (that would cause the attitudinal account to collapse into the motivational account). Instead, oughts are normative if they involve attitudes because the attitudes *are attitudes of approval and disapproval*, which are themselves normative notions<sup>351</sup>

### **Imperative Oughts**

Moral oughts are normative in virtue of being commands, and commands are normative because they are stated in the imperative mood<sup>352</sup>

These are all conceptions of *oughts without normative practical reasons of rationality*:

### **Oughts without Normative Practical Reasons or Rationality**

Oughts that are normative without being or entailing normative practical reasons of rationality or rational requirements<sup>353</sup>

I will consider these options in more detail shortly. Here I just want to make the point that the moral oughts that Desire-Independence mentions are consistent with many different metaethical theories of oughts. Therefore, we know that Desire-Independence is sufficiently inchoately formulated to count as a platitude. We also know that Desire-Independence all by itself cannot settle the case for moral error theory or indeed any other metaethical theory. So, for instance, moral rationalists who believe that moral reasons are reasons of rationality will emphasize the rationality-involving conception of morality's normativity that is consistent with but not entailed by Desire-Independence. But other metaethicists who aren't rationalists in this sense won't insist on this, and this is also consistent with Desire-Independence.

A third comment about Desire-Independence is that it is consistent with moral oughts being *dependent* on various aspects of the agent and her situation. For instance, it might be true of all of us that we morally ought to give to famine relief if we have a salary in excess of the minimum required to live a decent life in a certain

---

<sup>351</sup> Parfit (2012 Vol2: 268); Russell (2011: 209).

<sup>352</sup> Parfit (2012 Vol2: 268); Copp (2007: 257-8).

<sup>353</sup> Kieseewetter (2012: 466). I add specifically 'normative practical reasons' because Oughts Without Normative Practical Reasons or Rationality is consistent with oughts being analysable in terms of moral or legal 'reasons' understood as considerations that count in favour of X-ing without that implying truths about rationality.

country, regardless of whether we desire to give to famine relief and regardless of what we come to believe after having reflected on the matter long enough, *et cetera*. Nevertheless, the truth of this moral ought claim does not entail that we all morally *have to give to famine relief*. That depends on whether one earns enough.

This ends my discussion of Desire-Independence. I will now consider some other platitudes about the nature and status of morality.

A second platitude about the status of morality is:

### **Overriding**

Sometimes if an agent morally ought to  $\phi$  then that is simply what she ought to do<sup>354</sup>

According to Overriding, sometimes what agents morally ought to do completely determines what they ought to do all things considered. Try reading Overriding with ‘refrain from causing an innocent child to undergo a prolonged and excruciatingly painful death’ for ‘ $\phi$ ’ (cf. §1.1). In this case, most of us think, doing what we morally ought to do is *simply* what we ought to do full stop. Overriding is consistent with many morally relevant actions, such as not scoffing at beggars, for which it is not true that we simply ought not to do that full stop. Perhaps scoffing at a beggar is the only way to preserve an important work of art (the assumption here is that the badness of scoffing at beggars is outweighed by the (non-moral, aesthetic) goodness of preserving an important work of art). For all that Overriding says is that some moral oughts are all-things-considered oughts. Before moving on, let me note that Overriding is also consistent with different theoretical accounts of the way in which moral oughts are normative.

The third platitude concerns the way in which we moralize:

### **Moral Discussion**

When we moralize we exhibit a “disposition to persist in arguing, to continue giving reasons, to refuse to accept that convergence is unavailable, to be disappointed when argument runs out.”<sup>355</sup>

---

<sup>354</sup> McNaughton (1988: 114); Parfit (2012 Vol1: 146).

<sup>355</sup> Lillehammer (2004: 100); also see Kalderon (2005: Chapter 1); Brink (1986: 23). Lillehammer infers from this that we have “*prima facie* evidence for a conception of moral thought on which satisfaction of the convergence claim constitutes part of what it is for moral judgments to be true”, where ‘convergence claim’ is his version of the Non-Negotiable Commitment Claim and states that moral discourse couldn’t be moral discourse if we don’t connect it to the convergence of all rational beings on the matters (cf. the inter-subjectivity option from §3.3).

Consider Abe and Bert. They are arguing over the wrongness or otherwise of stealing. It is platitudinously true that—unless they get tired, unless some other (important) things interfere with their exchange of ideas, and unless a stalemate or impasse becomes obvious, *et cetera*—they will tend to continue to argue their position. It is also true that, under the same conditions, they will tend to hold their ground if no good reason to change their view is forthcoming.<sup>356</sup> They also refuse to accept that there isn't a right answer that they will converge on if they talk about the issue long enough (assuming, perhaps, sufficient capacity for rational deliberation, open-mindedness, and the like).<sup>357</sup> Moreover, if it is explained to them that beauty may be in the eye of the beholder and that morality might be like that, both will protest at this idea.<sup>358</sup> Finally, the folk will be disappointed when argument runs out, such as when Abe announces that his fundamental moral principle is simply that he can do whatever he likes. This third platitude is also consistent with various theoretical interpretations of the way in which moral oughts are normative.

Fourthly, we can note as a platitude:

<b>Phenomenology</b>	Morality feels objective to us—morality seems to have an external, inescapable authority <sup>359</sup>
----------------------	---

It is widely known, and has been widely noted, that morality has a distinctive, objectivist feel. Moral demands feel to us as if we cannot escape from them—as if they are simply laid down on us for us to respond to. Whether there is something that corresponds to this feeling remains to be seen, of course. It also remains to be seen whether taking this platitude on board requires postulating an account of moral oughts according to which they equate to rational oughts, or whether perhaps a rule or imperatival conception of moral oughts suffices. But that this feeling exists is beyond question.

---

<sup>356</sup> Enoch (2011a: 9).

<sup>357</sup> Smith (1994).

<sup>358</sup> Bloomfield (2007: vii).

<sup>359</sup> Kirchin (2003); Horgan and Timmons (2005, 2008a, 2008b); Mandelbaum (1955); Mackie (1977); Joyce (2009).

### 5.3 The Moral Metaphysical Entailment Claim

I will now argue that with these four platitudes about the status of morality we get an entailment to a particular kind of moral normativity that is stable under various mutations of this list of platitudes (that is, the entailment remains in place even when we add further—at least contender—platitudes to the list). I first argue, against relativists like Harman, that morality's purported normativity is *categorical* rather than *hypothetical* (§5.3.1). Recall the following distinction (cf. §2.3.1). (1) morality has categorical normative force *without* this entailing anything about the rationality of the agents who recognize this (morality is *categorically applicable*). (2) morality has categorical normative force and this *does* entail something about the rationality of the agents who recognize this—viz., this entails that agents are irrational when they ignore morality's categorical normativity (morality is *categorically reason-giving*). Harman's position is that morality is categorical in neither sense. His position is coherent and is inconsistent with my own, and so although he is not my most formidable opponent (most metaethicists believe that (1) is true), I should argue against him.

Next I argue against my most important opponents (§5.3.2). These include those naturalist cognitivist realists who agree with me (and disagree with Harman) that morality is categorically *applicable* but disagree with me that morality is categorically *reason-giving*. It is also in this section that I explain why the combined results of §5.3.1-2 justifies a Metaphysical Entailment Claim.

#### 5.3.1 Categorical Moral Normativity

In my attempt to argue that morality is categorically applicable (rather than merely hypothetically applicable, in which case the ascription of moral obligations would be sensitive to addressees' desires and social roles), I start by discussing how Desire-Independence and then Overriding, Moral Discussion and Phenomenology favour the categorical applicability view. I then add various other (at least contender) platitudes of and thought-experiments about moral discourse to the picture and show that the commitment to the categorical applicability of morality's normativity remains in place.<sup>360</sup>

---

<sup>360</sup> I say (at least) 'contender' platitudes as it is not entirely clear whether these claims are in fact platitudes, as will be explained shortly.

The thesis that morality is categorically applicable provides a clear fit with Desire-Independence:

**Desire-Independence**

If an agent morally ought to  $\phi$ , then that agent morally ought to  $\phi$  regardless of whether she feels like  $\phi$ -ing (at the time) and regardless of whether  $\phi$ -ing would further any of her goals even if she thought about  $\phi$ -ing for a long time

Suppose Desire-Independence is true. Suppose, that is, that if an agent morally ought to  $\phi$ , then that agent morally ought to  $\phi$  regardless of whether she feels like  $\phi$ -ing (at the time) and regardless of whether  $\phi$ -ing would further any of her goals even if she thought about  $\phi$ -ing for a long time. We might conjecture that this is so because moral oughts are categorically applicable. After all, what it means for an ought to be categorically applicable is for that ought to constrain or suggest an action independently of whether the agent that may perform the action desires to perform it or can be gotten to desire to perform it by being shown that the action accords with her ends or desires.<sup>361</sup> So saying that morality's normativity is categorically applicable explains Desire-Independence very nicely.

Next consider:

**Overriding**

Sometimes if an agent morally ought to  $\phi$  then that is simply what she ought to do

I surmise that saying that morality's oughts are categorically applicable *and* very robust explains this datum. For if morality's normativity isn't categorically applicable then agents can wiggle around their desires and in so doing escape their moral obligation. But if agents can escape their moral obligation in this way then Overriding can't be true. Now of course, if morality's normativity doesn't *also* at least sometimes override other kinds of normativity by being the most robust kind of normativity then Overriding will be false. So in accounting for this platitude we have to make this further assumption as well. However, this doesn't mean that it isn't necessary in accounting for Overriding that morality's normativity is categorically applicable. It is just not necessary *and* sufficient.

Now consider Moral Discussion:

---

<sup>361</sup> Wallace (MS: 5).

### **Moral Discussion**

When we moralize we exhibit a “disposition to persist in arguing, to continue giving reasons, to refuse to accept that convergence is unavailable, to be disappointed when argument runs out”

Suppose it is *not* true that moral oughts are categorically applicable (i.e., suppose that it is not true that moral oughts apply to everyone regardless of their desires), and suppose we all know that. Had this scenario been actual then we would expect to have observed that people stopped arguing for their moral beliefs a long time ago. After all, if you can escape a moral judgment from applying to you by changing your desires then a disposition to persist in arguing and to refuse to accept that convergence on the issue is unavailable is futile and wastes energy. In that case, all one’s interlocutor has to do to *legitimately* convince you that what he does isn’t wrong is show that he desires to perform that action (and perhaps that he cannot be gotten to desire otherwise). Continuing to browbeat him with moral reasoning is a waste of energy. But, we in fact do have this disposition to continue to argue, and we certainly think that it *isn’t legitimate* in moral discussions to cite outlandish and idiosyncratic desires. So it seems that moral oughts apply categorically.

Finally consider:

### **Phenomenology**

Morality feels objective to us—morality seems to have an external, inescapable authority

The thesis that morality’s normativity is categorically applicable also explains why Phenomenology is platitudinously true. Often what we ought to do depends on our desires. If I desire a chocolate ice cream on a hot summer’s day then what I ought to do is go to the shop and buy a chocolate ice cream. The authority of this practical ought is neither external nor inescapable. But with morality what we ought to do bears no such connection to our desires. The authority of moral oughts is external, or, precisely, they are categorically applicable. Now in order to explain Phenomenology we also have to assume that morality is robustly normative—perhaps by being rationally authoritative—but that doesn’t mean that morality’s categorical applicability isn’t necessary. It is—it just not necessary and sufficient.

Overall then, the idea that morality is categorically applicable provides an extremely nice fit with each of the four platitudes that I mentioned. I will now argue

that this remains the case even if we wield in (alleged) data from ordinary moral thought and talk that deny these platitudes.

Consider this from Harman:

My moral relativism is a soberly logical thesis—a thesis about logical form ... We make inner judgments about a person only if we suppose that he is capable of being motivated by the relevant moral considerations<sup>362</sup>

Harman denies Desire-Independence on account of the ‘logic’ inherent to moral discourse: ascriptions of wrongness, obligations and the like (‘inner judgments’) only make sense against the backdrop of knowledge that the person one is ascribing thin moral notions to can be moved by these ascriptions. Just as it makes no sense to say that a boat morally ought not to steal (because boats cannot be moved by the relevant considerations), it makes no sense to say that a psychopath or another character we know cannot be moved by moral considerations morally ought not to steal. Whether one morally ought to do something depends on what one desires or perhaps on what one can be gotten to desire under ideal circumstances of reflection. This would get us:

**Desire-Dependence**

It is not the case that if an agent morally ought to  $\phi$ , then that agent morally ought to  $\phi$  regardless of whether she feels like  $\phi$ -ing (at the time) and regardless of whether  $\phi$ -ing would further any of her goals even if she thought about  $\phi$ -ing for a long time

The question is, does this threaten our conclusion from before that morality’s normativity is categorically applicable? It doesn’t. First, there is the general countervailing consideration that

The many critical howls that greeted [Harman’s] claim are evidence that [Desire-Independence] was regarded by his audience as a conceptual truth<sup>363</sup>

Given this consideration it seems safe to assume that Harman is simply wrong to deny morality’s normativity categorical applicability.

---

<sup>362</sup> Harman (1975: 3-4), where “inner judgments include judgments in which we say that someone should or ought to have done something or that someone was right or wrong to have done something” (p5); also see Shafer-Landau (2005: 118n4).

<sup>363</sup> Shafer-Landau (2005: 118n4).

I won't however exclusively rely on this. What I will rely on is a version of Joyce's Jack the Ripper thought-experiment as discussed in §2.3.1 (suitably altered to avoid the problems with Joyce's original case discussed there).

But let me first explain the dialectic. I have just argued that there is a fair number of platitudes that support the view that morality's oughts are categorically applicable. But there is also countervailing evidence (Harman's Desire-Dependence). My argument that this doesn't jeopardize my claim that morality's normativity is categorically applicable has two prongs. First, there is evidence that Harman's view isn't true, let alone *platitudinously* true (see the quote above). From this it would follow that we cannot use Desire-Dependence *as a platitude*, and therefore that we cannot add it to my original list of platitudes to see whether this results in the view that morality's normativity isn't categorically applicable. But I also said that I won't rely exclusively on this idea. So here is a second, additional response. There are various *thought-experiments* that also entail that Desire-Dependence isn't (platitudinously) true. But that's not because the Jack the Ripper cases that I will mention are themselves platitudes. They aren't; they are thought-experiments with which philosophers can argue that certain theses, including a thesis like Desire-Dependence, are false. With Desire-Dependence down as false, it can no longer be used as a platitude that problematizes my argument so far, and I can conclude that the remaining platitudes about morality entail that its purported normativity is categorically applicable.

Consider, then, the following as a thought-experiment:

**Jack the Ripper-Deliberation**

Jack the Ripper can 'realize' that killing is wrong (i.e., fully believe that killing is wrong, have epistemic access to that belief, know that there are no epistemic defeaters around, etc.) and yet be without error if he ignores the wrongness of stealing<sup>364</sup>

What is our intuition about Jack the Ripper-Deliberation? Our intuition is that Jack the Ripper *has* made a mistake. My claim is that this intuition is best explained on the basis of the claim that morality is categorically applicable. After all, had morality been hypothetically applicable then Jack may be without error if he ignores

---

<sup>364</sup> Jack the Ripper is one of the world's most famous serial killers. I what follows I assume that Jack desires to kill for pleasure, and that he doesn't desire not to kill.



the wrongness of stealing. For then the normativity of the wrongness of stealing would be tied to some of Jack's desires, and we know that his desires favour killing over not-killing (or at least so we can assume for the sake of argument). Therefore, Harman is wrong when he says that inner moral judgments (involving deontic moral terms) can only be sensibly applied to agents with whom we have some kind of agreement in desires. Jack the Ripper-Deliberation shows that we *do* make inner judgments about him. Indeed, even when we learn that he is a psychopath—even when we learn that he was so deprived that he couldn't possibly be gotten to be motivated by the 'relevant moral considerations'.<sup>365</sup> Of course, the question of the *nature* of the mistake is still wide open. We haven't yet decided whether Jack ignored a rule or whether he was also irrational. No matter. For all we intended to do here is establish that morality's normativity is categorically *applicable*.

We can describe a similar case that goes not in terms of making a mistake but in terms of feeling guilt:

**Jack the Ripper-Guilt**

Jack the Ripper can 'realize' that killing is wrong (i.e., 'really' believe it, have epistemic access to that belief, know that there are no epistemic defeaters around etc.) and yet proceed to kill (or decide or settle on killing) without feeling or indeed having to feel guilt

My reasoning here is analogous to my reasoning in the Jack the Ripper-Deliberation case. First, our intuition is that Jack the Ripper *should* feel guilt. Second, we explain this on the basis of the view that morality is categorically applicable. After all, had it been part of the 'logic' of inner moral judgments that they only apply to agents who share a similar motivational structure with us, then it wouldn't be the case that Jack ought to feel (moral) guilt for what he did (recall our assumption that he has fundamentally different desires from us). He could just cite his different desires and that'd be that. But we do believe that Jack ought to feel guilt. So morality's normativity must be categorically applicable.

In addition to Jack the Ripper-Deliberation and Jack the Ripper-Guilt we can consider:

---

<sup>365</sup> Wallace (MS: 5) writes that "it is natural to put the point in the language of discretion, saying that [Jack doesn't] have the ... liberty to reject the claims of [morality]."

### **Jack the Ripper-Critique**

Jack the Ripper can be rightly criticized for killing<sup>366</sup>

Some philosophers believe that any contender moral system that fails to make sense of any kind of criticizability whatsoever fails to be a *moral* system.<sup>367</sup> I argue that they are right. At the risk of sounding like a broken record, we should start with our intuition, which is that our critique of Jack the Ripper is justified. Second, we explain this on the basis of the view that morality is categorically applicable. After all, had it been part of the ‘logic’ of moral judgments that they only apply to agents who share a similar motivational structure with us, then it would *not* be the case that we can legitimately critique Jack for what he did. For recall our assumption that he has fundamentally different desires from us; Jack can cite those desires, and therewith escape our moral criticism. But we believe that we *can* legitimately critique him. So morality’s normativity must be categorically applicable.

Each of the Jack the Ripper cases that I have just discussed supports the view that Desire-Dependence can’t be true. Indeed, it is only categorically applicable moral normativity and the desire-*independence* of moral oughts that can explain these thought-experiments. This suggests that Desire-Dependence is false, and therefore that the four platitudes we discussed earlier support the view that morality’s normativity is categorically applicable.<sup>368</sup>

### **5.3.2 Categorical Moral Normativity in the Rationality-Involving Sense**

So morality’s normativity is *categorically* rather than *hypothetically* applicable. But recall that this is consistent with various interpretations of the *nature* of morality’s normativity. Moral oughts can either be normative in virtue of being or entailing oughts of rationality or else in virtue of being or entailing motivation, rules, attitudes, or by being expressible by imperatives. What I will now argue is that morality’s purported normativity must be understood in terms of rational oughts—moral oughts equate to reasons that are normative reasons in the standard sense of

---

<sup>366</sup> Cf. Jackson and Pettit (1995: 29).

<sup>367</sup> Copp (2007: 2-6); Sinnott-Armstrong (2008: 92).

<sup>368</sup> One explanatory hypothesis for why Harman nevertheless undermines Desire-Independence in favour of Desire-Dependence is that he believed that the only reasons we have are desire-based reasons (Harman 1975: 9), and he might have realized, before Mackie’s (1977) and Joyce’s (2001), that without a rejection of Desire-Independence this would entail an error theory of ordinary moral discourse.

reasons of practical rationality, or they equate to rational requirements (or perhaps a combination of these).

I will present an argument for this that takes the form of a dilemma for different interpretations of my Jack the Ripper cases. The Jack the Ripper cases, recall, are not platitudes. They are philosophical thought-experiments designed to check the soundness of my argument that the best interpretation of the actual platitudes—which admit of exceptions, various interpretations and can be denied on theoretical grounds—is that they entail that morality’s purported normativity is categorically applicable *as well as* normative in the rationality-involving way. Earlier I argued that the Jack the Ripper cases show that the various platitudes about the normativity of morality unequivocally support the view that morality’s purported normativity is categorically applicable. Now I will argue that the same thought experiments also show that morality cannot be normative without being normative in virtue of being related to rationality. This methodology is in line with the idea that a metaphysical entailment claim depends on what interpretation of morality’s purported normativity provides the best fit with the data that we have—including both platitudes and thought-experiments.

Here, then, is the dilemma:

**Dilemma**

We think that what Jack the Ripper did was wrong and that he made a mistake/should feel guilt/can rightly be criticized for his behaviour. This is so *either* because he culpably ignored whatever it is that *morally speaks against* killing without this entailing that his mistake is a rational mistake *or* because in making a moral mistake he also necessarily made a rational mistake.

On the first horn of Dilemma, the opponent of error theory must argue that moral normativity without rationality can explain our intuitions about the Jack the Ripper cases (these will be intuitions that come in addition to the ones we’ve already considered). I argue that this impossible, so that the second horn must be embraced. But on the second horn of the dilemma, the opponent of error theory must accept my claim that moral oughts are rational oughts.

Let’s consider the case that can be made for the first horn of the dilemma. Recall our four options for conceiving of morality’s normativity without rationality:

**Rule-Involving Oughts**

Moral oughts are normative in virtue of being rules, and rules are normative in virtue of distinguishing between what is *correct* and *incorrect* or what is *allowed* and *disallowed*

<b>Motivational Oughts</b>	Moral oughts are normative in virtue of entailing actual or possible motivation, such that when a putative ought, known by an agent, doesn't actually or possibly motivate, it fails to be an ought
<b>Attitudinal Oughts</b>	Moral oughts are normative in virtue of involving attitudes of approval or disapproval to our own or other people's actions. Attitudes of approval or disapproval are often motivationally efficacious, but this feature isn't what guarantees the relevant oughts' normativity (that would cause the attitudinal account to collapse into the motivational account). Instead, oughts are normative if they involve attitudes because the attitudes <i>are attitudes of approval and disapproval</i> , which are themselves normative notions
<b>Imperative Oughts</b>	Moral oughts are normative in virtue of being commands, and commands are normative because they are stated in the imperative mood

*Rule-Involving Oughts.* Can we explain what we want to explain about the Jack the Ripper cases when we conceive of morality's normativity as consisting of rules? Let's consider the case for this claim.

Start by taking a close look at the rules of chess. In chess it is a rule that you can only castle if you haven't yet moved your king. So, if you want to play *chess*, you shouldn't 'castle' after you've moved your king. But, of course, you could decide to play *schmess* and 'schcastle' whenever you want, and thus perform the move that is characteristic of castling after having moved your king.<sup>369</sup> So we have at least the following choice: playing chess and playing schmess. Surely each game is normative *in the sense that each game contains rules that tell potential players which moves are and are not allowed in the game*. However, the truth of the claim that it is *chess rather than schmess* that one should play *to begin with* depends on your *choice or desire*. You should play chess if you want to play chess and you should play schmess if you want to play schmess. What therefore explains the normativity of following the rules of *chess* is a *desire* to play chess rather than *schmess*. It is a rule in chess that you can castle only when you haven't moved your king yet, but this rule has normative force over you because and only because you desire to play *chess*.

---

<sup>369</sup> Enoch (2011b).

I am willing, at least *arguendo*, to accept that the normativity of the ‘castle’-rule can be fully reduced to the relation that obtains between my desire to play chess and the truth of the claim that you can only play *chess* when you adhere to the rule that you can’t castle after you’ve moved your king.<sup>370</sup> This probably also gets us a naturalistically respectable account of the truth of the claim that you ought not to castle when you’ve already moved your king. So we have a naturalistically respectable explanation of the ought claim that you should not castle if you’ve already moved your king (if you want to play chess). So far so good.

The question that concerns us here though is whether we can transpose this account of normativity as it pertains to the rules of chess to the normativity of morality. For what we want to have explained is our intuition that Jack the Ripper made a mistake when he decides to kill his victims and proceeds to do it (and that he should feel guilt and that we can rightfully criticize his behaviour). But here we find the problem for the opponent of error theory wishing to argue that it is not essential to moral discourse that moral oughts are rational oughts; and that, in particular, moral oughts can be understood as involving rules only. For let’s transpose the rule-conception of the oughts of chess to the oughts of morality. So let’s assume that we can have a naturalistically respectable explanation of the ought claim that Jack ought not to kill (*if he wants to play the ‘moral game’*—i.e., if he desires to heed to morality’s demands). This is all well as far as it goes, but we should note that the addition of the phrase ‘if he wants to play the moral game’ entails a rejection of our earlier established claim that morality’s normativity is categorically applicable. After all, on this conception of the oughts of morality, *if Jack doesn’t want to play the moral game then he is free to kill*, just like someone not wishing to play the chess game could play schmess instead and castle after having moved his king. But we think that morality’s normativity is categorically applicable, and hence that this kind of weaselling with one’s desires to escape moral obligations isn’t allowed. The problem is that the rule conception of oughts, as I argued earlier, depends on the obtaining of certain desires but that for *morality’s* oughts, this can’t be true. So we can’t use the rule-conception if we want to make sense of our intuition about the kind of mistake that Jack the Ripper makes.

---

<sup>370</sup> Olson (2009a, 2009b, MS); cf. Fitzpatrick (2008: 179-83).

The same holds, *mutatis mutandis*, for our intuitions that Jack should feel guilt and that our critique of him is legitimate. For if we accept the rule-conception interpretation of the normativity of morality, thus including the fact that it is desire-based then we get the result that Jack doesn't have to feel guilt for stealing when he forms a desire not to play the moral game. Perhaps Jack want to play the schmoral game, according to which killing is at least sometimes perfectly permissible if one wants to get pleasure from it, as is the case for Jack. But our intuitions are that Jack *does* have to feel guilt, and that it is not up to him to choose the moral over the schmoral game. So the rule-conception account, although it might work for the oughts of chess, cannot be used for the oughts of morality.

And the same again holds for Jack the Ripper-Critique. If we accept the rule-conception interpretation of morality's normativity then we get the result that we cannot rightfully criticize Jack's behaviour. For Jack, we assume, simply desires to kill for pleasure, in which case he can decide to play the schmoral rather than the moral game. And the rule-conception account of morality's normativity, relying as it does on the obtaining of our addressee's desires to heed to morality's demands, doesn't allow us to critique Jack. After all, he plays the schmoral game because he desires to do that, and we play the moral game because we desire to do that. But we most adamantly think that Jack cannot so decide. So the rule-conception account of morality's normativity is inadequate. It can only work if we assume that morality's normativity isn't categorically applicable—but we cannot assume that.

My opponents could object to this as follows. They could argue that Jack's mistake is a *contingent matter*. It is not simply and straightforwardly in virtue of being irrational that Jack wrongly kills. Instead it is in virtue of very reliable, but contingent, connections with Jack's concerns and that of this society and the content of the moral code that he makes a mistake in killing.<sup>371</sup> This objection will take some time to explain, and although I take it very seriously, I will answer it only in §5.4 below. So with this objection temporarily out of sight, I conclude that the rule-conception of morality's normativity fails to capture what is distinctive about the way in which morality is normative.

*Motivational Oughts.* Motivational oughts fare no better. According to the motivational conception of the normativity of a moral ought, moral oughts are

---

<sup>371</sup> Brink (1986); Copp (1995, 2007).

normative in virtue of entailing actual or possible motivation, such that when a putative ought, known by an agent, doesn't actually or possibly motivate it fails to be an ought. The problem with the motivational conception of normativity too lies with its ability to account for the categorical applicability of morality's normativity. For on the motivational conception of normativity, something can be normative if and only if it can motivate. But that precisely goes against the view that the normativity of morality applies to all of use *regardless* of what we desire (assuming, as is standard in current metaethical literature, that desires are motivationally efficacious mental states). Therefore, Motivational Oughts should be put aside.

*Attitudinal Oughts.* Attitudinal oughts do seem to be consistent with categoricity. For according to the attitudinal conception of normativity, an action is normative in virtue of being the object of an attitude of approval or disapproval. This attitude can be the attitude of one person directed at another person and her (possible) actions. I can express disapproval of you not giving to famine relief, and then for the true ought claim that you should give to famine relief to be normative it doesn't matter what *you* desire.

I will now argue that in fact this doesn't hold. Consider, once more, Jack the Ripper-Deliberation:

**Jack the Ripper-Deliberation**

Jack the Ripper can 'realize' that killing is wrong (i.e., 'really' believe it, have epistemic access to that belief, know that there are no epistemic defeaters around, etc.) and yet be without error if he ignores the wrongness of stealing

The question we must ask is this: is Attitudinal Oughts consistent with Jack the Ripper-Deliberation and our result from the previous sub-section that morality's normativity is categorically applicable? The answer is 'no'. For if all it takes for moral oughts to be normative is to consist of possible actions that are the object of attitudes of approval and disapproval then it will be *legitimate* for Jack to ask why he should play the moral rather than the schmoral game. Most people, let us assume, think that killing is not permissible, whereas Jack thinks that it is. Why is it that Jack has to take the fact that most people have a different opinion from him into account in deciding what to do and think? Why isn't it rather the case that *the other people* need to take the fact that *Jack* thinks that killing is permissible into account in deciding what to do and think? Given the symmetry here and the lack of further

resources on the part of the friend of attitudinal oughts, we have no reason to go either way on this matter. But then we can't explain our intuition that it is Jack who has to heed to our attitudes rather than the other way around. Again this account of the way in which moral oughts are normative can't respect morality's categorical applicability.

Advocates of the attitudinal account of normativity could object to this that we should idealize on information and reasoning in defining the relevant attitudes. They could argue that is not just the case that the attitudes that people as a matter of contingent fact happen to hold determine what is morally right and wrong. Instead, it is the *rationality* of these attitudes that determines which actions are right and wrong; viz., those actions that are the object of attitudes that it is *rational* to hold. But if we insist on rationality to this extent then what is doing the philosophical work in the attitudinal account is rationality, not the attitudes. Focussing on rationality to this extent vindicates the thesis of this chapter; viz., that you can't have moral normativity in the absence of (categorically applicable) rational constraints. So this objection, if anything, plays into my hands.

*Imperative Oughts.* Imperative oughts also seem to be consistent with the categorical applicability of morality's oughts. If oughts can be normative solely in virtue of being expressible by linguistic utterances in the imperative mood, then, on the obviously true assumption that we can express claims in the imperative mood without regard for desires, attitudinal oughts are categorically applicable.

But I want to challenge the idea that we can explain our intuition that Jack ought to play the moral rather than the schmoral game merely on the basis of the story that moral oughts are normative in virtue of being expressible in sentences in the imperative mood. If consisting of rules or being the object of attitudes is not enough, then being expressible as commands won't be enough either. For it is possible to place next to the demand against killing a demand *to kill*, or an expression of the permissibility of killing. But then, given morality's categorical applicability, how shall we explain Jack's mistake when he accepts that killing is permissible if we can't appeal to rationality? All we have again is a stalemate between on the one hand an imperative to steal and on the other hand an imperative to refrain from stealing. But if that is all we have then we can't explain why *Jack* made the mistake, and not us, when he decides to kill whereas we decide not to kill.



But if we can't explain this intuition then the imperatival conception of morality's normativity won't work either.

Now consider the rationality-involving sense of normativity. Why does this conception of the normativity of morality fare better? The answer is that the rules of rationality are inescapable in a way that merely moral rules (or imperatives, objects of attitudes *et cetera*) aren't. So suppose that we say that morality is normative in virtue of consisting of the rules of *rationality*. Clearly, the rules of rationality apply to all agents regardless of what they desire—the rules of rationality are the rules of *rationality proper*, not of particular *loci* of rationality such as Jack the Ripper, president Obama, or whomever. The key question is whether Jack the Ripper can intelligibly ask why he should follow the rules of rationality in the same way as he could intelligibly ask whether he should follow the moral rules, whether he should care about the attitudes of other people, whether he should care about some imperatives but not others, *et cetera*. This is the key question because if we can say that this question cannot be intelligibly asked then conceiving of morality's normativity in the rationality-involving way allows us to explain our intuitions about the Jack the Ripper cases.

The answer is that, indeed, the rules of rationality cannot be intelligibly ignored. Consider Joyce's example of an unwilling captive in Roman times who is thrown into the arena and forced to choose between fighting for his life or getting killed.<sup>372</sup> He is told that he can fight in whatever way he pleases, but that he is not allowed to throw sand in his opponent's eyes. Within no time the unwilling captive finds himself in a very bad position and the only way to stay alive is to throw sand in his opponent's eyes. The unwilling captive can legitimately question the rule not to throw sand in his opponent's eyes. He can say 'so what?' throw the sand, and save his life. In this story, because the alleged rules of gladiatorial combat have almost nothing to do with the captive (the only relation between the rules and him is that they are *forced upon him*), the rule not to throw sand can be legitimately ignored.

This contrasts with cases in which the rules of an activity are much more closely tied to the agent. Take, indeed, the rules of rationality as they govern desire and belief-formation. In a sense these rules aren't 'the agent's' either. The agent was never *consulted* about which rules are to be accepted just like the unwilling gladiator

---

<sup>372</sup> Joyce (2001: 34-6).

was never consulted about which rules of combat are to be accepted. But in another good sense of ‘the agent’s’ the rules of rationality are clearly ‘the agent’s’. For agents like us necessarily have desires and beliefs, and the requirements that govern them cannot be shrugged off in the very same way that the rules of gladiatorial combat can be shrugged off (or attitudes, or the rules of a ‘moral’ or ‘schmoral’ system, or the fact that a rule can be communicated with a sentence in the imperatival mood, *et cetera*). Agents don’t have to be gladiators—certainly not against their will—but agents have no choice but to have beliefs and desires. The rules of rationality cannot be legitimately ignored. But we know from our platitudes and thought-experiments that moral rules too are such that they can’t be legitimately ignored. Therefore, tying moral rules to rational rules allows us to vindicate morality’s stringent normativity, at least in principle. And therefore, we should understand the way in which moral oughts are normative as involving rational oughts. (Unfortunately for realist theories of morality, my argument in the next chapter is that there are no rational rules with recognisably moral content).

Let’s take stock. The discussion in the previous section has allowed us to claim that moral oughts are *categorically applicable*. The discussion in the present section has allowed us to claim that the categorical oughts of morality are normative in virtue of a connection with rationality. Taken together, this allows us to formulate the Metaphysical Entailment Argument:

**Metaphysical Entailment Argument**

Premise One:	There are various claims that are contenders for platitudinous truth, including Desire-Independence, Overriding, Moral Discussion, Moral Motivation and perhaps Desire-Dependence.
Premise Two:	The best interpretation of the state of affairs described in Premise One is that morality has to be categorically normative in the rationality involving way.
Premise Three:	If Premise Two, then propositions about morality metaphysically entail propositions about categorical moral normativity in the rationality-involving sense even though this is not built into the (application conditions of) moral concepts that we use to express these propositions.
Conclusion:	Therefore, propositions about morality metaphysically entail propositions about categorical moral normativity in the rationality-involving sense even though this is not built into the (application conditions of) moral concepts that we use to express these propositions. [From Premise One, Premise Two, Premise Three]

The argument is clearly valid. Premise One was our subject in §5.2 and Premise Two was our subject in §5.3. So for this argument's soundness all we need is the truth of Premise Three. I include Premise Three and indeed the Metaphysical Entailment Argument because I initially, in Chapter 3, explained that metaphysical entailment relations are in the first instance relations between propositions. But all that Premise Three really is is another way of expressing Premise Two. So if the move from Premise One and Premise Two to Conclusion works then adding Premise Three to it to accommodate earlier terms of discussion will work as well.

Hence my conclusion. It is essential to moral discourse as we presently have it that there exists categorical moral normativity in the rationality-involving sense. Unless, that is, the objections to the extent that we can in fact do with a less demanding, non-rationality-involving account of morality's normativity I bracketed earlier and will now discuss work. But, as I will now argue, they don't work.

## 5.4 Doing With Less

My conclusion so far is that it is essential to moral discourse that there exists categorical normativity in the rationality-involving sense, as this is the only way to explain our intuitions about the Jack the Ripper cases and other platitudinously true claims about moral discourse. I've argued for this conclusion without taking into account the *content* of morality—I only talked about platitudes and thought-experiments about the *status* and *function* of morality. This, it could be argued, amounts to begging the question against those naturalists who believe that by taking morality's content into account they can explain these aspects of moral discourse without thereby making claims about a necessary relation between rationality and morality. It is this objection that I will now engage with.

Consider:

### **Content**

[Human maintenance] and flourishing probably consist in necessary conditions for survival, other needs associated with basic well-being, wants of various sorts, and distinctively human capacities. People, actions, policies, states of affairs, etc. will bear good-making moral properties just insofar as they contribute to the satisfaction of these needs, wants and capacities ... [and] there is reason to think that moral facts will at least typically provide agents with reasons for action. Everyone has reason to

promote his own well-being, and everyone has reason to promote the well-being of others at least to the extent that his own well-being is tied up with theirs<sup>373</sup>

The argumentative strategy of my opponent here is to underplay on the hand the practical role that moral facts can reasonably be expected to play and to emphasize on the other hand the practical role that these facts *can* play.<sup>374</sup> In particular, she insists that telling a convincing story about how moral demands are closely tied to things that are very important to all of us can get her a convincing explanation of the Jack the Ripper intuitions. And this holds, she argues, even though Jack the Ripper is constitutionally incapable of changing in ways that bring his desires more in line with the important things in life. What is important in a human life? Content gives the example of human flourishing. Given that we know that human beings are such that they flourish only if they give due consideration to their own and other people's well-being, form friendships (and perhaps participate in politics, engage in life-long learning, have satisfactory jobs, *et cetera*), it makes sense to say that there is reason of rationality for everyone like us to engage in these activities—at least in the vast majority of cases and in most circumstances. In particular, it is not necessary, claims this kind of naturalist, to suppose that someone like Jack the Ripper can be shown to be irrational every time he acts immorally. What is important is that we can explain the reliable, but contingent, connection between the content of morality and its normativity (and motivational force).

The fact that the content-naturalist is happy with a reliable but contingent connection between morality and its normative force is highly problematic. Take the case of Jack the Ripper considering causing an innocent child to undergo a prolonged and excruciatingly painful death. Again, the kind of naturalist I'm discussing would argue that Jack always makes a mistake when he kills, namely a moral one, but that there are cases, perhaps this one, where such a mistake lacks rational authority. The question is whether this gets the naturalist an account of our intuitions about the nature of Jack's mistake that is at least roughly equally good as that of the error theorist. This is the relevant question because if it does then my error theory becomes rationally arbitrary.

---

<sup>373</sup> Brink (1984: 122); also see Brink (1986); Railton (1986, 1989: 154); Copp (1995, 2007); Finlay (2011: 546-7).

<sup>374</sup> Sturgeon (2006: 110). The 'practical role' of moral facts concerns both their normativity and their motivational force.

The answer, though, is that it doesn't. For our intuitions about the Jack the Ripper cases are that, *most emphatically*, Jack makes an important mistake when he sets out to kill innocent children. Admittedly, content naturalism can capture *a lot*, and can explain *a lot*, about this intuition. It is probably safe to assume that many agents, even those relevantly similar to Jack, will be swayed by the arguments of the content-naturalist. Moreover, the content naturalism-story provides a lot of depth to the explanation of what's wrong with Jack's behaviour—indeed, the kind of explanatory depth that we expect a successful explanation of what's wrong with Jack's behaviour to provide. The explanatory depth I'm referring to includes reflections on the content of the moral rules, the working of society and the nature of human flourishing. Finally, the content naturalist does this in an intuitively plausible way and with fairly minimal assumptions.

The problem for the content-naturalist is that the folk will *never* accept the thought that Jack can be acting fully rationally and yet immorally. This was the message of §5.3 (and especially §5.3.2), and looking at the *content* of morality as opposed to *the way in which it is normative* (e.g., by involving rules or by being expressible in sentences in the imperatival mood, *et cetera*) isn't going to change this verdict. The folk won't accept the thought that Jack can be fully *rational* when he kills and that 'all' that is wrong with him is that he is immoral. The folk want morality to be more *oomphy*. They want morality to be much more closely tied to what affects Jack's practical deliberation. The folk will *never* accept the thought that although there are various considerations that affect Jack's practical deliberation—such as means-end considerations and considerations pertaining to prudence more generally—morality is something that agents can completely ignore, at least on certain occasions and in certain circumstances, without being irrational.<sup>375</sup> True, these are quasi-empirical data that stand in need of further empirical corroboration, but on the assumption that these data are correct (and it is my quasi-empirical proposal that they are), it follows that the content-naturalist cannot explain our intuitions about Jack's behaviour as well as the moral rationalist who thinks that immoral behaviour entails irrational behaviour. The objection from Content doesn't work.

---

<sup>375</sup> On this point also see Copp (2012: 31-2).

Or does it? There is one further objection that we should deal with. Consider the thought that the folk will think that Jack the Ripper is *not* a full-fledged moral agent when he behaves in the way he does (killing and stealing at will). And so, the thought continues, the folk won't attach any importance to the kind of considerations I have been mustering in this chapter. After all, on this proposal, they think that Jack the Ripper falls outside of the scope of moral assessment. Or else Jack the Ripper is a sufficiently far out case that the folk are, perhaps on reflection, prepared to acknowledge that their intuitions about him might not be reliable.

Again my argument stands in need of further empirical corroboration. In line with my methodology of quasi-empirical metaethical research, I can honestly report that in my well-confirmed experience, even the big baddies in history, according to the folk, are subject to a robust sort of blame and have made a robust sort of mistake that outstrips the kind of blame and mistake that we can associate with a rule- or attitude-conception of morality's normativity (or any of the other, non-rationality involving ones). Indeed cases like that of Jack the Ripper come to mind. I will heed to the outcomes of proper empirical research and invite my readers to consider whether they include these kinds of characters in their moral assessment and what they think of them. But it is my own clear intuition that Jack the Ripper falls within the sphere of moral assessment. As far as the objection from Content goes, therefore, error theory and its commitment to moral rationalism is safe.

## **5.5 Conclusion and Preview**

I have argued that error theorists can get a working account of the Non-Negotiable Commitment Claim if they take this to be an instance of metaphysical entailment between moral propositions and propositions about categorical moral normativity understood in the rationality-involving sense. The remainder of the thesis argues that there is no categorical moral normativity thus understood, at least not in the actual world (Chapter 6). It also argues that objections thus far ignored don't hurt my error theory all that much (Chapter 7) and that a version of revolutionary cognitivism is the best response to this version of error theory (Chapter 8).

## Chapter 6

# Categorical Moral Normativity in the Rationality-Involving Sense

### 6.1 Introduction

In the previous chapter I argued that folk moral discourse carries a non-negotiable commitment to categorical moral normativity in the rationality-involving sense. Categorical moral normativity in the rationality-involving sense is the kind of normativity that moral facts would have to possess to *count* as moral facts.

One way to make this idea more precise is to consider the nature of practical rationality. Facts that exhibit categorical normativity in the rationality-involving sense must be taken appropriately into account in an agent's practical deliberation, irrespective of what the agent already desires to do or what his social roles happen to be. For practical reasoning is the type of activity that is sensitive to facts that are normative in the rationality-involving sense.<sup>376</sup> Agents who let the specific normative weight of facts with categorical normativity in the rationality-involving sense affect their practical deliberation to the required degree are taking those facts appropriately into account in deciding what to do. If fact F counts in favour of  $\phi$ -ing to degree d then agents who factor the fact that fact F counts in favour of  $\phi$ -ing to degree d into their deliberations about what to do are deliberating correctly. Contrastingly, agents who don't do this are not deliberating correctly.<sup>377</sup> The non-

---

<sup>376</sup> See Copp (2007: 270); Gibbard (1990: 61-80); Shafer-Landau (2003: 165); Jackson and Pettit (1995: 29).

<sup>377</sup> This is a somewhat simplified picture. For instance, deliberation isn't always necessary for successfully taking the normative weights of various facts into account. People might *automatically* be truthful without giving truthfulness any thought and this action might nevertheless be fully appropriate and morally laudable, especially when the relevant norms have been internalized by the agent (Copp 2010: 152). What is essential to morality is that agents take the normativity of moral considerations appropriately into consideration in thinking about what to do, think, and feel; whether they manage to do that by conscious deliberation or through earlier internalization of norms is a bit of a moot point.

negotiable commitment of moral discourse is to the truth of the thesis that there exist facts that are normative in this way and have recognizable moral content.

In arguing for this version of the Non-Negotiable Commitment Claim I am in broad agreement with existing moral error theories such as Joyce's and Olson's. As Copp puts it:

if morality has ... authoritative normativity, its having this property consists in the obtaining of the relevant specified relation between it and the norms of rationality<sup>378</sup>

But where I parted from Joyce and Olson is that I left it completely open whether or not normativity in the rationality-involving sense can be analysed without remainder, on the level of concepts or properties, in terms of reasons—a proposal which Joyce and Olson accept.<sup>379</sup> But normativity in the rationality-involving sense can involve more than reasons or perhaps doesn't involve reasons at all.<sup>380</sup> Indeed I will use the idea that the normativity of morality may also consist in *rational requirements*, which are also normative in the rationality-involving sense:

**Rational Requirements**

Requirements that govern the holding of various combinations of mental states, such as desires and beliefs, in abstraction from the question of whether we have reasons to have those attitudes in the first place<sup>381</sup>

It is one of my main goals in this chapter to argue that denying that normativity in the rationality-involving sense must be analysed without remainder, on the level of concepts or properties, in terms of reasons allows the error theorist to defend a more plausible version of error theory than has thus far believed to be possible.

---

<sup>378</sup>Copp (2007: 271).

<sup>379</sup> Or to be more precise, Joyce thinks that the normativity of morality must be analysed in terms of reasons (see §2.3.1 of this thesis) and Olson's currently published work commits him to it (although in his MS he writes that he now favours the view that the culprit of morality is to do with irreducible favouring-relations—see §2.4 of this thesis). For the view that normative claims are or involve claims about reasons see Scanlon (1998); Raz (1999).

<sup>380</sup> In what follows I will sometimes, when context allows me to, omit the qualification 'in the rationality-involving sense' to avoid endless repetition.

<sup>381</sup> Broome (1999). Also see Kolodny (2005); Parfit (2012: Vol. I, p36); Southwood (2008: 20); Brunero (2010: 31-2).



The structure of the chapter is as follows. I start by arguing that rational requirements exist because they are naturalistically respectable (§6.2). I then argue that rational requirements govern all sorts of (combinations) of desires and beliefs. For example, suppose Abe has an intrinsic desire to smoke, the belief that the only way to smoke is to buy a new pack of cigarettes in this very shop, but that Abe doesn't have the instrumental desire to buy a new pack of cigarettes in that shop. In that case Abe is flouting the following rational requirement (where RR is to be read as 'rationality requires that')<sup>382</sup>:

**Instrumental Principle (IP)\***

RR (you take the means to your ends)<sup>383</sup>

Indeed, I argue that the rational requirements that exist vindicate our platitudinously true beliefs about what is prudentially and epistemically rational or irrational (§6.3). This means that my account of normativity in the rationality-involving sense is independently plausible. Many things for which we want to have a vindicatory explanation—including our belief that the normative claim that Abe ought to buy a new pack of cigarettes is true—can be vindicated without violating a plausible metaphysical naturalism. For the vindicatory explanation invokes rational requirements like IP\* and these can be grounded in a naturalistic metaphysic, as argued in §6.2.

This also means that I accept that there exists *categorical* normativity in the rationality-involving sense. For instance, the Instrumental Principle is categorically normative in the rationality-involving sense as it is inescapably authoritative for agents with the capacity for rational agency regardless of whether or not they desire

---

<sup>382</sup> Smith (e.g., 2007) prefers 'reason requires that'. Kolodny (2008: 439) claims that there is a difference between the locutions 'reason requires that' and 'rationality requires that'. Southwood (2008: 20) only talks about requirements: the requirements of practical reason tell us how we ought to act, the requirements of theoretical reason tell us how we ought to believe, and the requirements of practical rationality tell us how we ought to reason. I will not take sides on the issue of which of these philosophers has latched on a better nomenclature and simply note that 'rationality requires that' I intend to refer to the kind of normativity that the rational requirements exhibit.

<sup>383</sup> I call this IP\* because this is a rough formulation of the principle; it will be made more precise later in the chapter.

to abide by it.<sup>384</sup> My account is committed to the truth of the claim that there are desires that agents rationally ought to have and where the ‘ought’ has categorical force. However, these oughts are also relative or sensitive to which other desires and beliefs the agent already has. Agents ought to desire to buy a new pack of cigarettes if they believe that buying a new pack of cigarettes is the only, or the best, way to satisfy their non-instrumental desire to smoke and if they have a non-instrumental desire to smoke.

In fact my picture of normativity in the rationality-involving sense is a bit more complicated than this. I will argue that in addition to true categorical ought claims that tell agents to adopt certain *instrumental* desires or beliefs *in light of other desires and beliefs they already have* (the Abe case), there are also some *intrinsic* desires and beliefs that agents ought to have where the ‘ought’ has categorical force *and where this claim isn’t made true by reference to other attitudes*. This includes the intrinsic desire to believe what one has sufficient evidence to believe. Unfortunately for the moral success theorist, there are no rational requirements that oblige us to adopt intrinsic or instrumental desires with recognizable moral content (§6.4). And looking at moral *reasons* rather than rational requirements won’t bring solace either (§6.5). Therefore, we should be local moral error theorists, believing that although whole rafts of normative claims in the rationality-involving sense are true, a subset of those claims—those that involve *morality*—are never true. This allows error theorists to avoid the Formulation Dilemma from §1.4: we can have a local moral error theory.

I call the view that there are lots of rational requirements, including requirements to adopt intrinsic desires with *non-moral* content, Proceduralism+:

**Proceduralism+**

Agents with the capacity for rational choice and action are bound by rational requirements to have certain desires and beliefs, not to have certain other desires and beliefs, and not to have certain combinations of desires and beliefs. But it is never the case that such agents are bound by rational requirements to have or not to have desires with recognisably morally relevant content, such as a desire not to steal.<sup>385</sup>

---

<sup>384</sup> Hampton (1998); Dreier (1997: 96).

<sup>385</sup> To save space I also sometimes write ‘recognisable moral content’ instead of ‘recognisably morally relevant content’ (even though this is not entirely accurate as

The ‘+’ in Proceduralism+ usually indicates that Proceduralism+ accepts that there are other, purely procedural norms in addition to the procedural norm which is expressed by the Instrumental Principle.<sup>386</sup> Procedural norms tell agents to adopt desires or beliefs by reference to the other desires and beliefs the agent already has. Substantive norms tell agents to adopt certain intrinsic desires without reference to the other desires and beliefs the agents already have. However, I will use the term Proceduralism+ to refer to a view that accepts both procedural norms other than the Instrumental Principle and substantive non-moral norms *but never substantive moral norms*. I do not intend to use this name in a proprietary way, however, and something like Substantivism- would work equally well (where the ‘-’ would indicate that there exist substantive norms albeit not with recognisable moral content).

Such is going to be my argument against categorical moral normativity in the rationality-involving sense. I close by summarizing my main findings and previewing what is to come (§6.6).

## 6.2 Grounding Procedural+ Normativity

In this section I argue that rational requirements are naturalistically respectable. For my own metaethical theorizing and that of the vast majority of my interlocutors, the issue of the naturalistic respectability of metaethical theories’ postulations is very important.<sup>387</sup> It is for this reason that I assume, at least for the purposes of this thesis, the importance of naturalistic respectability.

### 6.2.1 Functional Norms

Consider the following examples of rational requirements:

#### **Instrumental Principle (IP)**

RR (if you have an intrinsic desire that p and a belief that you can bring about p by bringing about q, then you have an

---

‘stealing’ is not a moral term, so that a desire not to steal is not as such a desire with moral content).

<sup>386</sup> Smith (2012a).

<sup>387</sup> Also see, e.g., Timmons (1999); Scanlon (2009); Miller (2013).

instrumental desire that you bring about q)<sup>388</sup>

### **Belief Consistency**

RR (you don't have logically inconsistent beliefs)<sup>389</sup>

Copp assures us that “realism about rational requirements ... is compatible with metaphysical naturalism.”<sup>390</sup> But how does this naturalisation of rational requirements work?

Some philosophers adopt the view that there is a limit to normative explanation; viz., the explanation of what it is in virtue of which certain facts are normative. For instance, Shafer-Landau argues that a defence of the kind of view that I am defending “always consists in offering replies to criticisms, and evidence far short of demonstrable proof as positive support.”<sup>391</sup> I however will offer positive

---

<sup>388</sup> Smith (2007: 282, 2009: 101, 104); Railton (1986); Joyce (2001: 54); Brink (1989: 63); Skorupski (2007: 76); Hubin (1999: 32; 2001: 459); Way (2012: 487-8); Setiya (2007: 649); Bratman (2009: 13); Darwall (1983: 9, 2001: 141-2); Railton (2006: 265); Joyce (2001: 54). Note three things about IP. I. IP can take entities other than just desires as its object of assessment, including ‘the will’, and can admit of slightly different formulations in both cases (Zangwill 2012: 354). II. The instrumental principle plausibly extends beyond productive to constitutive means (Korsgaard 1997: 215-6: 652n7; Nagel 1970: 51-2; Wedgwood 2011). III. ME is formulated as a wide-scope (WS) rather than a narrow-scope (NS) requirement because its normative force governs the entire conditional rather than just the consequent; schematically (where ‘O’ stands for ‘ought’): if you have end E  $\rightarrow$  O take the means to M (NS), O (if you have end E, then take the means to E) (WS). WS is more plausible than NS because NS allows for detaching of the consequent; simply by adopting whatever end it is that I fancy I can make it the case that I ought to take the means to that end (these would include the end of killing myself and that of killing others, which are implausible on prudential and other grounds). WS avoids this by leaving it open how one resolves the irrationality of holding end E and not the means to E; i.e., by leaving it open whether one drops the end or takes the means (so that in the killing case one can simply drop that end).

<sup>389</sup> Kolodny (2005: 510, 2008: 437); Southwood (2008: 10n4). We can add to these: Desire Coherence (RR: you don't have non-instrumental desires that fail to cohere with each other) (Smith (2004: 108); Intention Coherence (RR: you don't have intentions that fail to cohere with each other, such as an intention to X and an intention not to X) (Parfit (2012: Vol. I, p36); Southwood (2008: 10n4); Enkrasia (RR: you intend to do what you believe you ought to do) (See, e.g., Broome 2008: 98; Copp 2007: 272), to name just a few.

<sup>390</sup> Copp (2007: 314).

<sup>391</sup> Shafer-Landau (2003: 210).

support for Proceduralism+. I will argue for the *proper functioning account* of rational requirements and that this account entails their naturalistic respectability.

Consider functional kinds such as knives.<sup>392</sup> The metaphysically mundane fact that *knife* is a functional kind grounds the true normative claim ‘it ought to be the case that knives cut well’ (where the ought is not necessarily an all-things-considered ought—I will leave out this specification for reasons of space). For what a knife *is* is something whose *function* it is to cut well. This allows us to say that ‘it ought to be the case that knives cut well’ is an *efficient way of saying* that the evaluative claims such as ‘knives that cut serve their function better than knives that don’t’ and ‘knives that cut more efficiently function better than knives that cut less efficiently’ are true. And these evaluative claims *themselves* are not queer because they are *short-hand* for the descriptive claim that a given knife is higher up a list that reports to what extent, compared to others, it exhibits properties that abide by a given standard for knives.<sup>393</sup> Again, “no Moorean non-natural qualities are ... required to underwrite the truth of [these claims].”<sup>394</sup>

The proper functioning account shares relevant similarities with Copp’s pluralist teleological theory of normativity. That theory “accounts for normativity by reference to the idea that a normative system can have the “function” of ameliorating a problem of normative governance.”<sup>395</sup> Recall Mackie’s schmoral normativity from §2.2.4. This was a kind of normativity that doesn’t involve objective prescriptivity. With it Mackie tried to solve the problem of sociality, which was the following problem of normative governance. How can we best govern a complex system of interacting individuals that makes up our current society, the

---

<sup>392</sup> Thomson (2008); Smith (2010: 124-5). Nagel (1970: 18) anticipates arguments such as these when he writes that his “method is one of metaphysical ethics: moral and other practical requirements are grounded in a metaphysics of action, and finally in a metaphysics of the person. The more central and unavoidable is the conception of oneself on which the possibility of moral motivation can be shown to depend, the closer we will have come to demonstrating that the demands of ethics are inescapable.”

<sup>393</sup> Bloomfield (2001: 19, Chapters 3-4).

<sup>394</sup> Smith (2010: 124, also see his 2012b: 203, 2012a: 241-2, 2010: 125).

<sup>395</sup> Copp (2009a: 33; also see 2012: 38). The proper functioning account also has affinities with Aristotle (see his *Nicomachean Ethics*; Joyce 2006: 169; Casebeer 2003).

function of which, according to Mackie, is to guarantee that humans can live harmoniously and reap the benefits of social cooperation?

In this chapter we have a different, albeit similar, problem of normative governance. Here we look at our individual psychologies, not societies, and the problem concerns the complex system of interacting *attitudes* that makes up the human psychology and how it can perform its function. Call this not the problem of sociality but the *problem of psychology*. What the function of our psychologies is will be explained shortly. It is a moot point in the present context in which I only aim to explain that the proper functioning account of rational requirements is an instance of what Copp calls a teleological theory of normativity.

The proper functioning account should be distinguished from the *intrinsically valuable status account*, according to which what grounds the truth of normative claims such as the Instrumental Principle is that failing to obey a rational requirement involves somehow undermining or failing to *respect that status*.<sup>396</sup> For the proper functioning account doesn't mention the intrinsic value of obeying rational requirements. It only uses the idea that obeying rational requirements is the only way to perform one's function as a being with a psychology like ours.

Consider the claim that the psychology of an agent is a functional kind. For the psychology of an agent is something whose function it is *to produce knowledge (or true/reliable belief) and action*.<sup>397</sup> This means that psychologies, like knives, can be ordered according to how well they serve their function. This also means that we can ground the existence of rational norms in this description of the function of psychologies, just as we did with the norms governing the working of knives. Psychologies that work by adhering to IP and Belief Consistency are better placed to perform their function of producing true beliefs and actions than psychologies that don't. For instance, if you don't form the non-instrumental desire to transfer money to Oxfam's bank account when you intrinsically desire to 'give money' to Oxfam and know that transferring money to its bank account is a good (enough) way of 'giving money', then your psychology is failing as a functional kind. You won't be *acting well* or perhaps not acting *at all*. You are merely letting things happen to your

---

<sup>396</sup> Kolodny (2005: 544-7); Southwood (2008: 15).

<sup>397</sup> Smith (MS); Railton (1997: 64). For the point about producing knowledge also see Velleman (2000); for the point about producing action also see Bratman (1986).

body (or in this case, letting nothing happen to your body at all, because you are not transferring money). Again, saying that IP is true is just to say that this evaluative, but metaphysically mundane fact, is true.

We can tell a similar story that grounds rational requirements not in the functional profile of a human psychology as a whole but in the specific attitudes that are part of that psychology.<sup>398</sup> Zoom in on two essential elements in our psychologies: desires and beliefs. Take desires. It is a commonplace that desires are the states they are in virtue of their characteristic aims. They aim at their own satisfaction (compare: knives are what they are in virtue of their characteristic ‘aim’: they aim to cut). We can give this commonplace a functional interpretation: a desire is a disposition to realize its own content. Now desires manifest themselves in agents, and as such they need some background conditions to obtain for them to be able to perform their characteristic functions. On the one hand, desires need the agent to be able to move her body. On the other hand, since desires can be for anything at all, for a desire to realize its content it would have to combine with some representation of how the world is or needs to be changed by the agent. This means that, for a desire to be everything that it is supposed to be (i.e., for a desire to be able to display its functional behaviour), the desire would have to be such that

when an agent who has this desire in addition believes that its content can be realized by his performing some bodily movements, it combines with that belief in such a way that the agent is disposed to perform, and performs, that bodily movement. This is tantamount to saying that a desire functions optimally only if it conforms to the principle of instrumental rationality<sup>399</sup>

A similar story to that about IP and how it can be grounded in our psychologies can be told about Belief Consistency. Having a psychology that abides by this rational requirement will be helpful from the point of view of acquiring true

---

<sup>398</sup> Smith (2012b: 203-5); Kolodny (2008: 438). I assume that the folk psychological concepts of BELIEF and DESIRE that we use to pick out these mental states are not in error, and hence that, for instance, the desire theory of our mental states is false (Altham 1986). In the present context of a defence of moral error theory this is unproblematic, as many of my strongest opponents in the metaethical debate make the same assumption. Hence the assumption is something we all share and that I can argue *from*; it is not something I have to argue *for*.

<sup>399</sup> Smith (2012b: 205). “[It] is in the nature of ... desires to be subject to rational requirements” (Smith MS: 2).

beliefs and producing actions. If you have logically inconsistent beliefs then you can be sure that at least one of them is false. This contradicts our aim of having only true beliefs. This explains why we should abide by Belief Consistency.

Why is the function of our psychologies to believe true things and to produce actions that satisfy its desires? Because—and this much simply appears to be rock-bottom—that is what our psychology *do* and are *there for*. Our psychologies enable *us* to *do* things; our psychologies are not mere *loci* of events. A table is an example of a mere *locus* of events. There are external forces impeding on it, such as gravity, rain and sunlight (if the table is stood outside), and it can be moved by intelligent beings like ourselves (we can take the table inside to prevent it from getting wet or discoloured). But *we* are not mere *loci* of events in that sense. It is not as though there are only forces from outside of us, or forces from inside of us that we cannot control, that makes it the case that there are merely things happening to us. We can move about ourselves. An instance of the problem of normative governance therefore applies to us. Here we have a complex system of interacting attitudes that is the human psychology, including motivationally efficacious states; what are the rules that, when the agent accepts them, would best allow her to have a psychology that can optimally perform its function of getting the agent to move about in the world?<sup>400</sup>

There are problems with this account of the function of our psychologies, as there are problems with other aspects of my proper functioning account of rational requirements. I formulate and respond to them in the next sub-section.

### 6.2.2 Problems and Solutions

First, there is the objection that the error theorist can't simply assume that the function of our psychologies is to produce action and true beliefs *and just that*. What, for instance, about the suggestion that our psychologies exist, at least in part, to allow us to get along with other humans? If that were part of the functional profile of our psychologies then presumably we *can* ground rational requirements with categorical force and recognizable moral content on it. And from this a moral success theory would follow. I take this objection seriously, but postpone answering it until §6.4.2.

---

<sup>400</sup> Copp (2012: 39).



A second objection is that agents are seemingly able to simply shrug off the alleged normativity of rational requirements. For why would the fact that agents wouldn't have a 'properly functioning psychology' if they don't abide by IP and all the other rational requirements *matter to them*? Indeed, what's wrong with saying, 'fine, I won't have a properly functioning psychology if I don't take the means to my ends; but so what?'

Copp makes a similar objection to Thomson's functionalist view of human beings.<sup>401</sup> According to Thomson, being morally virtuous is the function of human beings. Copp complains that he can imagine Nietzscheans who, precisely, *don't* want to be morally virtuous and are still able to live their lives and indeed seemingly coherently so. This would be evidence against the claim that humans wouldn't be functioning *as humans* if they aren't being morally virtuous. Scanlon use a different example to make the point:

It may be true in a functionalist sense that we 'ought' to have the capacity to reproduce—that we are functionally defective if we lack this capacity. But these norms, and 'oughts', need have no normative force for an agent who recognizes them<sup>402</sup>

Likewise, it might be objected to my account that the rational requirements that it postulates have no *guaranteed* normative force for the agent who recognizes them.

However, we should draw an important distinction between *legitimate* and *illegitimate* 'so what?'-responses. Legitimate 'so what?'-responses are 'so what?'-responses that make sense. Recall Joyce's example of an unwilling captive in Roman times who is thrown into the arena and forced to choose between fighting for his life or getting killed.<sup>403</sup> He is told that he can fight in whatever way he pleases, but that he is not allowed to throw sand in his opponent's eyes. Within no time the unwilling captive finds himself in a very bad position and the only way to save his live is to throw sand in his opponent's eyes. The unwilling captive can legitimately question the rule not to throw sand in his opponent's eyes. He can say 'so what?', throw the sand, and save his life. In this story, because the alleged rules of gladiatorial combat have almost nothing to do with the captive (the only relation

---

<sup>401</sup> Copp (2011: 182); Thomson (2008).

<sup>402</sup> Scanlon (2007: 86-7).

<sup>403</sup> Joyce (2001: 34-6).

between the rules and him is that they are *forced upon him*), the ‘so what?’-response makes sense.

This contrasts with cases in which the rules of an activity are much more closely tied to the agent. Take, indeed, the rules of desire and belief-formation. In a way these rules aren’t ‘the agent’s’ either. The agent was never *consulted* about which rules are to be accepted; just like the unwilling gladiator was never consulted about which rules of combat are to be accepted. But in another good sense of ‘the agent’s’ the rational requirements that I’ve been exploring are clearly ‘the agent’s’. For agents like us necessarily have desires and beliefs, and the requirements that govern them cannot be shrugged off in the very same way that the rules of gladiatorial combat can be shrugged off. Agents don’t have to be gladiators—certainly not against their will—but agents have no choice but to have beliefs and desires. It doesn’t make sense to respond to the rational requirements by saying ‘so what?’ even though it does make sense to have that response when presented, unwillingly, with the rules of gladiatorial combat. So the objection that rational requirements, on the proper functioning theory of their normativity, can fail to be normative for agents who do not desire to abide by them doesn’t work.

A third objection to my view is that rational requirements are, in fact, not normative at all but only *appear* to be normative from the agent’s point of view. Some philosophers hold this view because they think that the only way to vindicate the robust sense in which rational requirements are normative is by showing that we can have a (substantive) *reason* to be rational, but that we have no such reason, and that we should therefore conclude that in fact rational requirements at most appear to be normative. Of course rational requirements could be normative in a more minimal way—they could be normative only in the imperative- or rule-involving sense. But, the objection goes, that wouldn’t be a vindication of the *robustness* of the normativity of rational requirements—that wouldn’t be a vindication of their normativity in the rationality-involving sense.

This is Kolodny’s ‘transparency account of the rational requirements’.<sup>404</sup> If Kolodny is right then I’m wrong in thinking that rational requirements are normative in the robust, rationality-involving sense. But Kolodny isn’t right. Kolodny claims that vindicating the normativity of rationality requires showing that we have a

---

<sup>404</sup> Kolodny (2005: 545-6); Southwood (2008: 13).

(substantive) reason to be rational, but he does this without considering the proper functioning account. So this claim begs the question against that account. For according to my account, vindicating the robust sense in which rational requirements are normative does not necessitate giving *reasons* to be rational. The only thing that is required is the kind of story about the function of our psychologies as I have told it so far. To simply assume that I'm wrong and that we need to be able to supply reasons for being rational is to beg the question against my account.

In addition to the transparency account of rational requirements there is the *epistemic* account of the normativity of rational requirements, as for instance defended by Setiya.<sup>405</sup> According to this view, and to take IP as our example, the ought that we invoke when we say that Abe ought to intend M (the means to his ends) is epistemic, not practical. The reason we should believe this account, according to Setiya, is that it is our last resort. The other views on the market don't even make sense.<sup>406</sup> But of course I claim that my own functionalist account of the norms of practical rationality does make sense and is plausible. And again this view is not even considered. In the absence of a focussed objection to my account, therefore, Setiya's critique too is question-begging.

The final problem is that I haven't been sufficiently clear about why my account is naturalistically respectable. What does it mean for an account of normativity in the rationality-involving sense to be naturalistically respectable? There is no shortage of suggestions. Parfit has two proposals. First, properties that are studied by the natural sciences count as natural properties. Second, properties that satisfy the Causal Criterion, according to which properties are real and natural only if they are causally efficacious, are natural properties.<sup>407</sup> G.E. Moore has used the following criteria: (i) natural but not non-natural properties can exist independently of any object instantiating them; (ii) natural but not non-natural properties exist in space and time; (iii) natural but not non-natural properties can be taken up in our hands or be separated from their bearers by scientific instruments.<sup>408</sup> Copp claims that a property is natural if and only if any synthetic proposition about

---

<sup>405</sup> Setiya (2007).

<sup>406</sup> Also see Schroeder (2009: 224).

<sup>407</sup> Parfit (2012 Vol2: 305-6); also see Enoch (2011a: 7, 159, 162, 177).

<sup>408</sup> Moore (1903: 41, 110, 124).

the instantiation of the property that can be known can be known only empirically, or by means of empirical observation and standard modes of inductive inference.<sup>409</sup> There are many more options, some evidently more plausible than others.

But which one should we accept? The problem is that without an adequate understanding of natural properties we can't really 'test' whether my account is naturalistically respectable. Moreover, in trying to settle on a criterion for naturalness we have to be wary of presupposing what we are trying to prove. If, prior to any investigation into whether normative properties can be natural we define all natural properties such that normative properties couldn't possibly meet these definitions then we are begging the question against the latter's existence.<sup>410</sup>

Enoch appeals in this context to the 'just too different intuition': normative properties are 'just too different' from all the natural properties in order to even qualify as a contender natural property.<sup>411</sup> To defend Enoch's suggestion: it is possible that for some fundamental debates, such as the debate about normativity and naturalness, intuitions carry a lot of argumentative weight and that we must ultimately rely on them in our philosophical accounts of the nature of reality. But at the same time it is also true that it would be *better* if we had more than just our intuitions to work with. So can we be more specific than Enoch's proposal and still somehow avoid having to settle the debate about the necessary and sufficient conditions of natural properties?

I propose a position about metaphysical naturalism that holds the middle between Enoch's suggestion that we can rely on intuitions and the very specific proposals of Parfit, Moore and Copp that specify necessary and sufficient conditions for what it takes for a property to be a natural property. We can work with examples of natural properties that are clearly recognizable as such, including the property of being an electron, being positively charged, being an Australian twenty-dollar bill or

---

<sup>409</sup> Copp (2007: 254).

<sup>410</sup> Miller (2003: 181).

<sup>411</sup> Enoch (2011a: 4, 107). Similarly, Shafer-Landau (2003: 80) writes that we "know [the natural] when we see it", much like we can recognize "the pornographic" when we see it. Cf. Joyce (2013). Enoch also accepts the more focussed proposal that "[f]acts and properties are natural if and only if they are of the kind the usual sciences invoke" (2011a: 103), but the just-too-different intuition seems to be doing most of the philosophical work in his thinking about the relation between normativity and naturalness.

the property a tree can have of being deciduous, and assess whether normative properties stray too much from these examples.<sup>412</sup> This stays rather close to Enoch's proposal, but also gives us a little bit more space to argue for our convictions. We can show that some normative properties, like that of being a rational requirement, should be included into our ontology because we can show that they are sufficiently similar to archetype natural properties. And we can show that some other normative properties, such as, perhaps, that of being a categorical reason, should not be included into our ontology because we can show that they stray too much from the archetype natural properties. This is not an ideal way to proceed—ideally we would have a full-blown account of natural properties that specifies their necessary and sufficient instantiation conditions—but the discussion of what makes natural properties natural would make my discussion unmanageable. So my argument is at least at this point subject to further corroboration.

Why don't rational requirements stray too far from archetype natural properties? Take, for instance, this piece of paper that has the property of being a £20 note. It has this property in virtue of complex facts about human agreements and economical facts *and in virtue of its function*. For a £20 note is the kind of thing you aren't allowed to use just by itself to pay for something that is more expensive than £20. It functions as a means to exchange goods and services with people in a more efficient way than trading those goods and services directly without the intervention of money, *et cetera*. So this piece of paper that is my £20 note is subject to various norms in virtue of its function. Likewise, this action of forming a desire to  $\phi$  has the property of being *required* by IP in virtue of complex facts about human psychologies *and their functions*. For this action is the kind of thing that you aren't allowed to perform if the right kinds of background conditions fail to obtain, such as the having of certain beliefs and intrinsic desires. The action is subject to norms in virtue of the function of the psychology that it is part of. There are important relevant similarities between being a £20 note and being a rational requirement. It is safe to assume that if the former is naturalistically respectable then so is the latter.

I will have more to say about the relation between naturalness and non-naturalness in due course. For now I close the topic and proceed to show that with

---

<sup>412</sup> Copp (2007: 254).

Proceduralism+ we can vindicate the truth of many normative claims in the rationality-involving sense, including some with categorical applicability.

### **6.3 Norms of Prudence and Epistemology**

In this section I list a number of platitudes about prudence and epistemology and I argue that they really are platitudes. I also argue that Proceduralism+ can vindicate these platitudes. If the project of this section is successful then Proceduralism+ is a plausible theory. This is important in the dialectic about error theory. Error theories are radical theories, and radical theories are often implausible. An error theory of all normativity would be more implausible than my local moral error theory which is compatible with a success theory about prudential and epistemic normativity in the rationality-involving sense. So this section blocks allegations of extreme implausibility as they beset other, more radical, error theories. The next two sections argue that a success theory of prudential and epistemic normativity in the rationality-involving sense is consistent with an error theory of moral normativity in the rationality-involving sense.

#### **6.3.1 Norms of Prudence**

Consider the following statements that I claim can be platitudes about prudence when we supply (at least some of) them with sufficiently determinate content (I present these statements here without content so that I can make some general remarks about them):

- (1) It doesn't make sense to have or fail to have certain intrinsic desires, unless one is in some such condition as exhaustion or lethargy or if one is under time-constraints.
- (2) The having or not having of certain instrumental or intrinsic desires doesn't make sense given certain background conditions that one is aware of, unless one is in some such condition as exhaustion or lethargy or if one is under time-constraints.
- (3) It doesn't make sense to fail to take into account all the important facts that one knows bear on one's decision as to what to instrumentally or intrinsically desire, unless one is in some such condition as exhaustion or lethargy or if one is under time-constraints.
- (4) It doesn't make sense to continuously hop between sets of intrinsic desires without being able to satisfy any of them, unless one is in some such condition as exhaustion or lethargy or if one is under time-constraints.<sup>413</sup>

---

<sup>413</sup> Henceforth I omit the qualification 'pace exhaustion, lethargy, and time-constraints' in all formulations of platitudes and norms; instead I'll take it as read.

Various comments about these platitudes are in order. First, I understand the location ‘it doesn’t make sense to’ as generically as possible.<sup>414</sup> That is, I use this locution as a means of reporting an intuitive sense in which things aren’t going as they should with the agents that have the (combinations of) mental states mentioned. What I want to avoid is stacking the deck in favour of a particular view about whether the kind of irrationality displayed is best understood in terms of irresponsiveness to reasons or in terms of something else.

Second, prudential considerations admit of vague boundaries in that it is often not entirely clear whether a consideration counts as a prudential one. Nagel equates prudence with practical foresight but also notes that it is more often equated with self-interest and that this latter understanding of prudence has a foundation in ordinary usage.<sup>415</sup> This latter understanding of prudence is what I shall have in mind throughout.

Third, the platitudes as I have them above seem to exhaust all the platitudes about prudence insofar as their *form* is concerned. It doesn’t make sense to have a desire, intrinsic or instrumental, with certain contents (platitudes 1 and 2). It also doesn’t make sense to fail to take into account all the important facts that one knows bear on one’s decision as to what to instrumentally or intrinsically desire (platitude 3). And once one has a set of desires it doesn’t make sense to change everything round all the time so that one cannot satisfy any of them (platitude 4). It doesn’t seem as though there are other *kinds* of considerations that I should take into account if I want to be prudent.

However, and this is the fourth point, we can’t be similarly exhaustive about the *contents* of these platitudes. There are many things that we could slot in as the objects of our desires in platitudes (1) and (2). And there are also many things that we might think are important considerations concerning the process of figuring out what our desires should be (platitude 3). Platitude (4) does not seem subject to this worry as it is purely formal (it doesn’t matter what kind of specific content for the desires we plug into it). I will argue later on that the fact that we can’t be completely exhaustive regarding the *content* of these platitudes is not a big problem for my claim that Proceduralism+ can account for all of the platitudes about prudence. The

---

<sup>414</sup> For this usage of ‘it doesn’t make sense’, see Railton (1997: 61).

<sup>415</sup> Nagel (1970: 36).

reason for this is that we have good reason to believe that my account of the platitudes that I do consider will extend to many other platitudes with different contents.

I will now list a number of platitudes *with content* that can be found in the literature. I put stars in front of them to distinguish them from their more general formulations (1)-(4):

- (\*1a) It doesn't make sense to intrinsically desire to have pain.<sup>416</sup>
- (\*1b) It doesn't make sense *not* to intrinsically desire to take medicine that will save one from lingering death, and restore one to perfect health.<sup>417</sup>
- (\*2a) It doesn't make sense to intrinsically desire to smoke, to believe that the only means to do that is to buy a new pack of cigarettes, but *not* to instrumentally desire to buy a new pack.<sup>418</sup>
- (\*2b) It doesn't make sense to believe that one should jump but fail to desire to jump.<sup>419</sup>
- (\*3) It doesn't make sense not to learn about what it involves to be a professional philosopher when considering going to graduate school.<sup>420</sup>
- (\*4) It doesn't make sense to desire to be a professional footballer one day, a professional philosopher the next day, a medic the third day, and so forth.<sup>421</sup>

One type of evidence that these platitudes really are platitudes is their ubiquitous recognition in the literature as such. A related type of evidence is that they wear their status as platitudes on their sleeves, so to speak. We find it difficult

---

<sup>416</sup> Parfit (1984: 212). Parfit's *datum* is actually formulated in a way that includes information about when the agony occurs (future Tuesdays) and has other specifications as well. But I think that all of these additional pieces of information about (\*1a) are inessential. After all, it doesn't seem to matter whether one desires pain on Tuesdays or on other days of the week—desiring pain *per se* doesn't make sense. Also see Heathwood (2011: 95-6); Finlay (2009b: 6); Street (2009: 273). Other examples to slot in (\*1a) include Hume's agent who prefers the destruction of the world to the scratching of his finger (1739-40: bk.2, pt.3, sec.3), Rawls's blades-of-grass counter (1971:432) and Gibbard's ideally coherent anorexic (1990: 171).

<sup>417</sup> Gert (2004: 41).

<sup>418</sup> Kolodny (2005: 509-10); Darwall (2006: 294); Finlay (2009a); Smith (2010, MS); Schroeder (2009); Brunero (2005); Rippon (2011); Dreier (1997: 89); Korsgaard (1997: 215, 2009: 1); Lillehammer (1999: 203).

<sup>419</sup> Smith (2009: 112, 2004b: 108); Kolodny (2005: 527); Parfit (2012: Vol. I, p36); Finlay (2009b: 4); Smith (1995: 118); Velleman (1996: 696-7); Parfit (1997: 101).

<sup>420</sup> Enoch (2011a: 88).

<sup>421</sup> Fehige (2001: 63).



to deny their truth, at least off the cuff. For instance, about something like (\*2a), R.E. Barnes writes that

a person could not coherently argue that she has no concern for taking the most efficient means to her ends. This kind of rationality is so foundational that rejecting it risks rejecting a necessary element of all practical reasoning<sup>422</sup>

However, a caveat applies. Consider for instance platitudes (\*1a) and (\*1b). It is sometimes claimed that pain and suffering are *good*.<sup>423</sup> This means that, although claims (\*1a) and (\*1b) clearly have a platitudinous standing of a sort, they are not the kinds of platitudes that are impossible to deny, such as the platitude that everything that is red is coloured, which is impossible to deny.<sup>424</sup>

I will now argue that that Proceduralism+ can provide a vindicatory explanation of these platitudes. Start with platitude (2):

- (2) The having or not having of certain instrumental or intrinsic desires doesn't make sense given certain background conditions that one is aware of, unless one is in some such condition as exhaustion or lethargy or if one is under time-constraints.

The Instrumental Principle (IP) can make good on this platitude. For according to IP, RR (if you have an intrinsic desire that p and a belief that you can bring about p by bringing about q, then you have an instrumental desire that you bring about q). On the plausible assumption that having a certain intrinsic background desire counts as a suitable background condition, IP says that if you believe you ought to pursue end E and if means M is the best or a good enough means to attain E then you ought to be motivated to pursue M, unless one is in some such condition as exhaustion or lethargy or if one is under time-constraints. That is, IP just *is* the second platitude couched in philosophical parlance.

How does this relate to (\*2a)-(\*2b)?

---

<sup>422</sup> Barnes (2012: 817). Also see Brink (1992: 11), who claims that “the link between an agent’s own interests and desires and his reasons for action seems intuitive.”

<sup>423</sup> Nietzsche often makes these claims; Parfit (2012 Vol2: 570).

<sup>424</sup> Shafer-Landau (2005); also see Jackson and Pettit (1995).

- (\*2a) It doesn't make sense to intrinsically desire to smoke, to believe that the only means to do that is to buy a new pack of cigarettes, but *not* to instrumentally desire to buy a new pack
- (\*2b) It doesn't make sense to believe that one should  $\phi$  but fail to desire to  $\phi$

(\*2a) is just an example of (2), and so because (2) can be accounted for my IP, so can (\*2a).

(\*2b) can be accounted for by:

**Enkrasia**      RR (you intend to do what you believe you ought to do)<sup>425</sup>

Enkrasia is another rational requirement like Belief Consistency and the Instrumental Principle. We have no reason to assume that it won't be naturalistically respectable, given that for it too we can tell a story about why psychologies that adhere to it are better able to perform their function than psychologies that don't. That explanation is that without Enkrasia, it will be harder, if not impossible, for agents to have their desires fulfilled, which is part of the function of their psychologies. And Enkrasia is simply (\*2b) in different words.

Let us now look at platitudes (1) and (3)-(4). First consider platitude (1); the platitude that the mere having of certain intrinsic desires, such as a desire for pain, doesn't make sense. Parfit argues that the Proceduralist+ cannot make good on this platitude. He thinks that she would have to accept state-given (i.e., mind-independent), objective reasons for attitudes. According to Parfit, what explains why it doesn't make sense for us to desire pain is that we have a *reason* not to desire pain—a reason, moreover, that is given by how things are rather than by what would satisfy our desires (hence the name state-given reasons).

I argue that in fact one of the suggestions that Parfit makes for the Proceduralist works.<sup>426</sup> The suggestion Parfit makes is that defenders of Proceduralism+ could explain platitude (1) by reference to the fact that pain stands in the way of achieving *whatever* you want to achieve. Given that desires are states

---

<sup>425</sup> See, e.g., Broome (2008: 98). Smith uses the slightly different formulation 'if you believe that you would want to  $\psi$  if you had a desire that conforms to all of the rational requirements, then you desire to  $\psi$ ' (2007: 281; 1994). Also see Copp (2007: 272).

<sup>426</sup> Parfit (2012, VolII: 85-9).

that aim at their own satisfaction, for desires to be able to perform and continue to perform that function it can't be the case that you desire pain. For that desire makes it impossible for you to continue to satisfy your other desires. After all, agents that are in pain are known to have difficulty satisfying their desires, whatever they may be. This gives us the following rational requirement:

**No Pain**

RR (if you have and exercise the capacity to act, then you desire not to be in pain)

And just as with Enkrasia we have no reason to think that No Pain isn't naturalistically respectable, given our background story about the function of our psychologies.

One problem for this suggestion is that there are (possible) agents who only desire pain and nothing else. Why should No Pain be a rational requirement *for them*? Off the cuff it doesn't seem like it should. So these possible agents are counterexamples to my account. Indeed, the existence of these counterexamples suggests that the rational requirements that I have just considered—Enkrasia, No Pain, *et cetera*—aren't *categorically* normative. Enkrasia and No Pain are normative for those agents who possess the right kinds of desires and beliefs. But for some other (possible) agents who do not have these desires and beliefs, Enkrasia and No Pain do not apply.

I don't think that this is much of a problem for my account. Platitudes are by their very nature amenable to rejection (in light of theoretical considerations) and come in various strengths. Above I gave the examples of the platitude that all red objects are coloured, which looks pretty unshakable, and Mackie's platitude, which can be doubted. Also recall that Nietzscheans can seemingly coherently deny platitudes (\*1a) and (\*1b). So the fact that I cannot *completely* deal with this platitude—i.e., that there are extreme cases which I cannot explain—is not a problem. Indeed, it may simply adequately reflect the (limited) contestability of this platitude. And indeed this platitude seems limited in scope. Almost everyone ought to abide by No Pain, except those probably only hypothetical characters who only desire pain. But that just seems to be the right result, not an objection to the view.

I have just explained how Proceduralism+ can provide a (sufficiently robust) vindictory explanation of platitude (\*1a). This leaves me to explain:

- (\*1b) It doesn't make sense *not* to intrinsically desire to take medicine that will save one from lingering death, and restore one to perfect health<sup>427</sup>

Here the reasoning is going to be the same as with (\*1a). Proceduralism+ explains platitude (\*1b) by reference to the fact that disease and death stand in the way of achieving *whatever* you want to achieve. Given that desires are states that aim at their own satisfaction, for desires to be able to perform and continue to perform that function it can't be the case that you fail to desire to be cured from a disease. For that desire makes it possible for you to continue to satisfy your other desires. After all, agents that are lethally ill are known to have difficulty satisfying their desires, whatever they may be. This gives us the following rational requirement:

<b>No Death</b>	RR (if you have and exercise the capacity to act, then you desire not to die)
-----------------	---

Of course, there will also be (possible) agents who only desire to be lethally ill and (almost) nothing else. But a story similar to which I just told in connection with agents who only desire pain and (almost) nothing else will apply here as well.

Platitudes (3) and (4) are still left. Start with platitude (4); the commonsensical observation that it doesn't make sense for agents to continue to hop back and forth between different life plans. Doing that makes it very hard, if not impossible, for you to satisfy any desire whatever. For if I am constantly occupied by changing my life plans and adjusting my beliefs and desires accordingly then I shall lack time and energy to actually do some desire-satisfaction. Moreover, desires that require more time to be satisfied—a desire to get a PhD in philosophy, for instance—are completely out of the picture if I change my life upside down all the time. This gives us the following rational requirement:

<b>Few Changes</b>	RR (if you have and exercise the capacity to act, then you desire not to change your life upside down too often)
--------------------	--

---

<sup>427</sup> Gert (2004: 41).

And again we have no reason not to assume that Few Changes is naturalistically respectable, given the availability of a story about the function of human psychologies.

Finally consider (3):

- (3) It doesn't make sense to fail to take into account all the important facts that one knows bear on one's decision as to what to instrumentally or intrinsically desire, unless one is in some such condition as exhaustion or lethargy or if one is under time-constraints.

If you don't acquire a reasonable amount of information about your circumstances, then it will be hard for you, if not impossible, to allow your desires to perform their function, which is to procure action with the aim of satisfying themselves. If I believe, falsely, that throwing my body in a fire gets me a holiday to Spain then after that I will neither have a holiday nor the means to satisfy any further desires at all. So we have:

**Information**

RR (if you have and exercise the capacity to act, then you desire to get a reasonable amount of information about the circumstances in which you act)

And at the risk of sounding like a broken record, again we have no reason not to assume that Information is naturalistically respectable, given the availability of a story about the function of our psychologies.

We can extend this line of reasoning to different platitudes that share their form with (1)-(4) but have different content. This is because it is plausible to believe that the move I've already used several times so far is *versatile*. The move was to point to rational requirements for which we have no reason to doubt that they can be grounded in the functional profile of human psychologies and that command agents not to frustrate their aims of realising their desires and attaining true beliefs. So for instance, recall:

- (\*2b) It doesn't make sense to believe that one should jump but fail to desire to jump.

I argued above that Enkrasia can deal with this platitudinously true statement. But we could also imagine a different way of giving content to (\*2b) such that Enkrasia can still provide a vindictory explanation of that platitude. So consider:

- (\*2b') It doesn't make sense to believe that one should duck but fail to desire to duck.

If Enkrasia vindicates (\*2b) it also vindicates (\*2b').

So to sum up, Proceduralism+ can account for the platitudes about prudence. I continue with the epistemological platitudes and will argue that Proceduralism+ can also provide a vindictory explanation for these platitudes. Moreover, I will argue that this consequence of my account pre-empts an important objection that many philosophers believe is devastating to error theory.

### 6.3.2 Norms of Epistemology

Consider the following claims that I argue have a platitudinously true standing (again, at least when we specify their content, which I again refrain from doing in the first instance because I first want to make some general points about them):

- (5) The having or not having of certain beliefs doesn't make sense given certain background conditions of evidence, unless one is in some such condition as exhaustion, lethargy, or unless one is under time-constraints.
- (6) It doesn't make sense to fail to take into account all the important considerations one knows bear on one's decision as to what to believe into consideration, unless one is in some such condition as exhaustion or lethargy or if one is under time-constraints.

Four comments apply. First, (5) and (6) are structurally similar to (2) and (4), respectively. It doesn't make sense to desire or not desire X given background conditions Y (i.e., platitude 2) is similar in structure to it doesn't make sense to *believe* or *not believe* X given background conditions Y (i.e., platitude 5). And 'it doesn't make sense not to take into account all the important considerations one knows to bear on one's decision as to what to believe or desire' simply integrates the contents of platitudes (3) and (6) into one.

Second, the first and fourth platitudes as we had them for prudential norms are absent. Consider platitude (1). Although it is platitudinously true that it doesn't make sense to have certain desires *per se* it isn't similarly true that it doesn't make

sense to believe something *per se*. For beliefs are *always responsive to the facts* whereas desires aren't (or at least need not be). And consider platitude (4). Although it is platitudinously true that it doesn't make sense to swap one's intrinsic desires round at such a speed that one cannot satisfy any of them, it is not similarly the case that it doesn't make sense to swap one's beliefs round at an identical speed. For one's beliefs are responsive to the facts, and if the facts change at a large speed (and if one is aware of that) then one simply has to adjust one's beliefs accordingly.

Third, evidence that these claims really are platitudes is simply their obviousness and again the fact that it is so often noted that they are platitudes (see the references I mention in the examples of (5)-(6) below).

Fourth, none of these platitudes is seemingly as easy to deny as the platitude that pain is bad, which, as we saw above, is denied by the Nietzschean.

Let's give platitudes (5)-(6) some content:

- (\*5a) It doesn't make sense to believe that the half-life of californium-253 is 17.82 days if all the facts point towards its half-life being 17.81 days and if one knows about these facts.
- (\*5b) It doesn't make sense to fail to believe that the half-life of californium-253 is 17.81 days if all the facts point towards the truth of that proposition and if one knows about these facts.<sup>428</sup>
- (\*6) It doesn't make sense to fail to take the (known) fact that it took californium-253's radioactivity to decay by 50% in 17.81 days into account in considering whether to believe that the half-life of californium-253 is 17.81 days.

I will now argue that Proceduralism+ can provide a vindictory explanation of these platitudes. Given that it is part of our proper functioning to believe only true propositions, the following rational requirement applies, for which we have no reason to assume that it isn't naturalistically respectable given the availability of the functional story of our psychologies:

<b>True</b>	RR (if you have the capacity for rational agency, then you desire only to believe those propositions for which you have sufficient evidence to believe that they are true). <sup>429</sup>
-------------	--

---

<sup>428</sup> Platitudes like (\*5a) and (\*5b), and the rational requirements corresponding to them, are discussed by Kolodny (2005: 521); Brunero (2010: 29); Railton (2003: 293); Stratton-Lake (2002: xxv-xxvi).

<sup>429</sup> Kolodny (2005); Brunero (2010).

True vindicates (5) and thus (\*5a) and (\*5b). For True forbids agents to form their beliefs on any other principle than that of sufficient evidence.

Regarding (6) and (\*6), the following RR applies:

**Information**

RR (if you have the capacity for rational agency, then you desire to get a reasonable amount of information about what you can and should believe)

Information provides a vindictory explanation of these platitudes as it says that agents would be violating genuine normativity when they don't take in sufficient information when they are considering what to believe. And again we have no reason not to assume that these rational requirements are naturalistically respectable.

Both True and Information are substantive rational requirements, and we are under *categorical* normative pressure to abide by them when it comes to our belief-forming processes. Hence my account of procedural normativity pre-empts a prominent objection to moral error theory, which is that its attempt to be a local error theory confined to *moral* discourse must collapse, implausibly, into a *global* error theory of epistemic and other kinds of categorical norms (this objection is sometimes called the Companions in Guilt Objection).<sup>430</sup> As I have just formulated this objection, it rests on the assumption that an error theory of categorical *epistemic* normativity in the rationality-involving sense is impossible or at least extremely implausible. But my formulation of error theory, because it accepts Proceduralism+, has no trouble dealing with this objection. For Proceduralism+ explains why there *is* categorical epistemic normativity in the rationality-involving sense and why there isn't categorical moral normativity in the rationality-involving sense (this will be argued in what follows).

This means that my error theory has a convincing reply to the formulation dilemma from §1.4:

**Formulation Dilemma**

---

<sup>430</sup> This objection is at the forefront of Cuneo's (2007a) book-length defence of moral realism. Also see Rowland (2013); Street (2009: 175-6); Lillehammer (2007); Streumer (forthcoming).



Error theorists can either formulate a *global* error theory via a broad meta-normative scepticism, or they can formulate a *local moral* error theory which leaves intact at least some normativity (and only finds, e.g., categorical moral reasons of rationality objectionable)

If error theorists embrace the first horn of this dilemma then the problem is that this renders their theory very implausible, as it is extremely hard to believe that there are no reasons for belief (or, to be more precise, that there is no categorical normativity in the rationality-involving sense that governs our belief-forming processes). But error theorists accepting the second horn of this dilemma must convince us that there is an important difference between categorical *moral* normativity in the rationality-involving sense and categorical *epistemic* normativity in the rationality-involving, and the problem with the second horn is that it is very hard to see what argument could deliver this result. The objection against error theory is that neither horn can be embraced and hence that error theory should be rejected. However, what I have argued in this sub-section is that, in fact, error theorists can provide a naturalistically respectable, vindictory explanation of epistemic normativity in the rationality-involving sense, and that the remainder of the thesis argues that this is still consistent with a rejection of categorical *moral* normativity in the rationality-involving sense. This allows error theorist to embrace the second horn of this dilemma and to avoid the objection that can be made on its basis.

The fact that Proceduralism+<sup>431</sup> vindicates categorical epistemic normativity in the rationality-implicating sense has further good-making features as well. One is that it can avoid the following, additional objection to error theory:

### **1. The Objection from Self-Defeat or Toothlessness**

Moral error theorists must either say that there are reasons to believe the error theory or that there aren't. If they say that there are such reasons, then moral error theory is self-defeating, for the property of being a (categorical) reason to believe shares all its essential characteristics with the property of being a (categorical) reason to act so that if the latter don't exist the former won't exist either. If they say that there are no reasons to believe the error theory then their theory is polemically toothless, for then it is not a rational mistake to reject moral error theory.<sup>431</sup>

---

<sup>431</sup> Cuneo (2007a: 117-8); Streumer (forthcoming: 13).

Proceduralism+ adopts the first horn of the dilemma, in a way, because it provides a vindictory explanation of how there can be genuine categorical normative pressure in the rationality-implying sense on agents to believe what they take themselves to have sufficient evidence to believe. Proceduralism+ adopts the first horn of the dilemma ‘in a way’ because it refuses to analyse epistemic normativity in terms of reasons, but that is a moot point in this context as nothing of importance hangs on this. Hence moral error theory is not polemically toothless.

A final objection that my error theory can deal with because it accepts Proceduralism+ is:

## **2. The Objection from the Normativity of Belief**

Beliefs are normative in the sense that a mental state *M* counts as a belief if and only if there is a reason for us to have *M* when there is (sufficient) evidence that the content of *M* is true. But if error theory about normativity as such is true then there are no mental states of which the foregoing is true. This means that there are no beliefs, which is a bad result for two reasons. First, it is hard to deny that there are beliefs. Second, this renders error theory self-contradictory: there are no beliefs and yet moral judgments express (false) beliefs.<sup>432</sup>

But my error theory, because it embraces Proceduralism+, denies that we should accept scepticism about normativity as such. It therefore avoids both the conclusion that there no beliefs and that error theory is self-contradictory. For the truth of Proceduralism+ entails that there *is* normativity in the rationality-involving sense to have a mental state *M* when there is (sufficient) evidence that the content of *M* is true.

So Proceduralism+ is plausible and saves my error theory from many implausible consequences and thus far it hasn’t vindicated distinctively moral norms. However, some philosophers have argued that Proceduralism+ actually entails that there are rational requirements to generate intrinsic or instrumental desires with recognizable moral content. The various ways of arguing for this entailment will be discussed in the next section, alongside my responses to these arguments.

---

<sup>432</sup> Shah (2010: 363-5); Streumer (forthcoming: 15); Cuneo (2007a: 121-2).

## 6.4 Procedural+ Categorical Moral Normativity?

In this section I argue that there are no substantive rational requirements with recognizable moral content, such as a requirement to muster a desire to refrain from stealing. In essence the argument for that claim is simple. The only naturalistically respectable rational requirements are those that can be grounded in the functional profile of human psychologies, which consists of the aims of acquiring true beliefs and producing action. This functional profile does not mention the aim of caring for or respecting other people or agents, and so any putative rational requirement that furthers *that* goal would have to come from somewhere else and is therefore metaphysically queer.

Various aspects of this argument can be questioned though. I first engage with the objection that there is an entailment from procedural+ norms to substantive norms that have moral content, which would allow categorical moral normativity in the rationality-involving sense to piggyback on the naturalistically respectable credentials of Proceduralism+. I argue that this entailment doesn't hold (§6.4.1). I also argue against the objection that the function of human psychologies should be described more broadly so as to include the aims of enabling agents to get along with other agents (§6.4.2). In §6.5 I argue against the suggestion that the normativity of morality should be understood in terms of *reasons* and can be vindicated in that way. This establishes moral error theory.

### 6.4.1 Do Procedural+ Rational Requirements entail Categorical Moral Rational Requirements?

In his most recent defence of a kind of moral realism of the rationalist variety, Michael Smith argues that procedural+ norms entail substantive norms with recognizable moral content.<sup>433</sup> To appreciate how this argument works first consider the belief formation processes that regularly occur in agents like us who possess the capacity to believe propositions on the basis of sufficient evidence. Suppose you are in the process of inferring that q from two premises: the premise that p and the premise that p implies q. You have arrived at the following point: you have settled that p, you are in the process of settling that p implies q and you are only

---

<sup>433</sup> Smith (2007: 285, 2009: 103, 2011: 354-60, 2012a, MS: 14-24); also see Markovits (2011: 277). I present Smith's argument with my own additions, clarifications, and interpretations.

anticipating performing the inference and forming the belief that *q*. A good question to ask is: what qualities do you have to have to believe that *q* on the basis of sufficient evidence at that very moment?

The answer is that you have to have various qualities. First, and perhaps trivially, you would have to have and exercise the capacity to believe propositions on the basis of sufficient evidence. But you also need to be able to rely on your past self who settled that *p*, to be vigilant at the moment in which you are settling that *p* implies *q*, and to be able to rely on your future self to draw the inference that *q*. In other words, the attitudes of reliance and vigilance are also required in addition to having a standing disposition to believe propositions on the basis of sufficient evidence.

Smith argues that what is required for vigilance in this sense is a desire not to allow other desires, such as those to believe things on the basis of wishful thinking, to interfere in one's reasoning. Hence what a rational agent should do is muster a desire with a particular content; namely, not to believe things on the basis of wishful thinking. And this is true in virtue of the fact that their psychologies are the kinds of things that work best if they believe things on the basis of sufficient evidence rather than on the basis of wishful thinking. So the following is a rational requirement, and indeed a substantive one:

<b>Vigilance</b>	RR (if you have and exercise the capacity to believe for reasons, then you desire not to allow wishful thinking to interfere in your deliberation)
------------------	--

What is required for reliance? Smith answers: a standing desire not to interfere with your future self's rational deliberation. After all, not having such a desire gets in the way of having a psychology that can perform its function of acquiring true beliefs. This gets us the following, again substantive rational requirement:

<b>Reliance</b>	RR (if you have and exercise the capacity to believe for reasons, then you desire not to interfere with your future self's deliberation)
-----------------	--

So far we have a kind of reasoning that is perfectly at home with the general tenets of Proceduralism+. Although we have categorical epistemic normativity in the

rationality-involving sense, we don't have recognizably *moral* norms. The kind of error theory that I defend can therefore accept everything that Smith has said so far.

The problem for my error theory though is that this reasoning has been argued to entail that there exist substantive *moral* rational requirements. For if what Reliance is meant to protect is *rational* deliberation, or its possibility, then it would be *rationally arbitrary* to apply Reliance only to the instantiation of the possibility for rational reflection *in oneself* but not to instantiations of the possibility for rational reflection *in other people*. And rational agents don't make such arbitrary decisions.<sup>434</sup> Therefore, rational agents should not only accept Reliance and Vigilance but also (just focussing on Reliance from now on to keep the discussion manageable):

<b>Reliance-Others</b>	RR (if you have and exercise the capacity to believe for reasons, then you desire not to interfere with your future self's or anyone else's deliberation)
------------------------	---

And Reliance-Others looks an awful lot like a moral principle; don't do to others what one wouldn't want to be done unto oneself. Or quite simply: don't hurt others.

Reliance and Reliance-Others are 'negative' principles in the sense that they only tell the agent what *not* to do. But they have 'positive' equivalents as well, such that the latter is derivable from the former in a way similar to that in which Reliance-Others was derivable from Reliance:

<b>Reliance-Positive</b>	RR (if you have and exercise the capacity to believe for reasons, then you desire to <i>positively</i> aid your future self's deliberation)
--------------------------	---

<b>Reliance-Others-Positive</b>	RR (if you have and exercise the capacity to believe for reasons, then you desire to <i>positively</i> aid your future self's or anyone else's deliberation)
---------------------------------	--

And Reliance-Others-Positive is also easily identifiable as a rational requirement with recognisable moral content, much in line with principles of beneficence that

---

<sup>434</sup> Smith (MS: 18). Nagel (1970: 114) speaks of 'practical solipsism' in this regard. Also see Copp (2007: Chapter 8); Joyce (2006: 197).

require or recommend agents to help others.<sup>435</sup> Call this version of the argument from entailment *Smith's Argument from Entailment*.

In response to Smith's Argument from Entailment, note a distinction that is not currently made in the literature. Distinguish between desires one should have on pain of irrationality that derive from the Instrumental Principle, Belief Consistency, Vigilance and Reliance on the one hand and desires one should have on pain of irrationality that derive from Reliance-Others and Reliance-Others-Positive on the other hand. The difference is that the first group of rational requirements can be grounded in *individual psychologies* (or the attitudes that constitute them, in the way explained in §6.2.1) whereas the second group rational requirements *cannot* be grounded in the functional characterization of *individual psychologies* (or the attitudes that constitute them, in the way explained in §6.2.1). After all, a requirement not to interfere with other agents' capacity for rational deliberation or theoretical reasoning, or indeed a requirement to positively help other agents with that capacity, *outstrips* what is required for having a working *individual* psychology, whose function is to produce true beliefs and action. But if rational requirements like Reliance-Others and Reliance-Others-Positive cannot be grounded in the functional characterization of individual desires then they aren't naturalistically respectable. But then it is not rationally 'arbitrary' to make a distinction between the two groups of rational requirements. For rational agents don't accept the existence of rational requirements that cannot be grounded in a naturalistically respectable metaphysic.<sup>436</sup>

---

<sup>435</sup> Defending the stronger positive theses isn't strictly speaking necessary—the weaker theses, which seem easier to defend will, all by itself, significantly compromise my error theory. I include a discussion of the stronger theses to show the versatility of this kind of objection to error theory.

<sup>436</sup> One way out for the success theorist, in addition to those that I will discuss (and reject) shortly, is to argue that such other naturalistically respectable phenomena as collective decision-making processes and groups can ground naturalistically respectable rational requirements or reasons that have recognisable moral content. In response, error theorists could attack the naturalistic respectability of group- or we-intentions, but even if error theorists leave this part of the objection intact two big questions remain. First, can the mere naturalistic respectability of a phenomenon ground true normative claims? It seems like it can't—we cannot make true normative claims about the property of being green *as such*. I argued that one way to bridge the gap is to invoke function-talk. But—and this is the second question—what would the *function* of group-intentions be? These questions need to be answered before the success theorist can make her case. Moreover, error theorists can rely on eliminative views about group-intentions according to which everything

My opponents have various rejoinders available to them. Parfit, for instance, starts by claiming that it would be irrational for individual agents to favour their own current selves over their later selves (or the other way round).<sup>437</sup> It would be irrational to take a drug that makes one feel great in the present time but that one knows will lethally damage one's liver in the long run. But if that's true, argues Parfit, then it would be *inconsistent* for rational agents to favour their own selves over other agents. After all, argues Parfit, there is *no relevant difference* between later selves and other agents. So we get rational requirements like Reliance (to help our later selves) *and* Reliance-Others (to help others) again. Call this *Parfit's Argument from Entailment*.

This version of the Argument from Entailment, apart from making some interesting additions to the earlier observation that it would be rationally arbitrary or *ad hoc* to accept Reliance but not Reliance-Others, adds nothing of importance to Smith's Argument from Entailment. In particular, it does nothing to assuage my worry that rational requirements like Reliance-Others are not naturalistically respectable. So again it is not rationally arbitrary to reject Reliance-Others and to accept Reliance because the former isn't naturalistically respectable.<sup>438</sup>

---

that needs to be explained about group-intentions can be explained just by using intentions of individual agents. But if that is so then what we are left with are individual psychologies and the rational requirements that can be grounded in *those*. But these don't leave room for requirements with recognisable moral content. Overall then, although this kind of objection to error theory first needs to be worked out before we can assess its viability fairly, it does already seem to be the case that error theorists don't have a lot to fear from it.

<sup>437</sup> Parfit (1984: §55).

<sup>438</sup> This debate mirrors to an extent the debate between the Nagel of his 1970 *The Possibility of Altruism* and his critics. Nagel writes "Just as there are rational requirements on thought, there are rational requirements on action, and altruism is one of them" (1970: 3). Also, "Nagel sees a parallel between intertemporal and interpersonal distribution of benefits and harms. He argues for agent-neutrality or altruism by analogy with prudence. Just as the interests of an agent's *future* self provide him with reasons for action *now*, so too ... *others'* interests provide *him* with reasons for action." (Brink 1992: 12). But the critique is that "Rational egoism assumes that *sacrifice requires compensation*, that is, that an agent has reason to make a sacrifice, say to benefit another, if and only if the agent receives some sufficient benefit in return ... [b]ut if sacrifice requires compensation, prudence and altruism must be importantly disanalogous. For, in the prudential case, I am compensated for a sacrifice of my present interests in favour of my greater future interests; these future interests are mine ... [b]ut interpersonal compensation is not automatic; benefactor and beneficiary are distinct. If the independence assumption is

We can also imagine someone arguing as follows. Call this the *Entailment Argument from Mistakes*. At least by hypothesis, she may start her argument, it is part of the function of our psychologies to produce true beliefs. But we know that individual agents who have the capacity for rational reflection often make mistakes. We also know that if agents with the capacity for rational reflection work together they are much less likely to make mistakes. So it would be rational for agents with the capacity for rational deliberation to work together, for the benefit of each individual agent. And to achieve such cooperation, adherence to such rational requirements as Reliance-Others is required. Therefore, such requirements as Reliance-Others *are* requirements for agents like us. Importantly, Reliance-Others has now been given grounding in individual psychologies of agents. For respecting this rational requirement *better enables* agents to achieve a psychology that can perform its function; viz., that of producing true beliefs. So it seems that the other-regarding (and therefore recognizable moral) rational requirements are in fact naturalistically respectable.

In response, error theorists can put pressure on at least one element of this argument. For there is the possibility of arguing that when individual agents *do* abide by all the rational requirements like True and Information they simply won't believe what there isn't sufficient evidence to believe. And in that case they will no longer need to respect rationality as it is instantiated in other agents. Take some of the well-known mistakes that people make when they are engaged in theoretical or practical reasoning. Depending on how certain problems are framed people are likely to affirm the consequent, to ignore relevant information, *et cetera*. Working together with other agents with the capacity for rational deliberation will decrease the possibility of making these mistakes. But so will taking the time to internalise

---

correct, the interest of other selves, however great, are not *ipso facto* interests of mine. Unless there is some connection between my interests and those of others ... I am not compensated when I sacrifice my interests ... for those of others. But then justified concern for my own future does not itself establish [or entail] justified concern for others" (Brink 1992: 12). The difference between my account of this kind of debate as presented in the text and the account Brink presents is the reference to the 'sacrifice requires compensation-thesis' has dropped out of my account in favour of what we may call the 'norms are all internal to the agent's own head (and have to be helpful to what's going on in there)'-thesis, although both Brink and I put emphasis on individuals in a sense. Also see Shafer-Landau (2003:199).



the rational requirements, grounded in one's psychology in the usual way, not to affirm the consequent and not to ignore relevant information. If this is right then as far as the goal of acquiring true beliefs is concerned, individual agents do not strictly speaking *need* other agents. True, agents may desire to be able to converse with other agents, and this would make it the case that they are, under pressure of the Instrumentalist Principle, rationally required to muster a desire not to kill other agents. But agents who do not have this desire and simply aim to ensure that their psychologies are in line with the rational requirements that there are won't be rationally required to refrain from killing other agents—they can perform their function all by themselves. But then a rational requirement like Reliance-Others can't get a grounding in the functional profile of individual psychologies *as such*—the rational authority of such requirement is desire-dependent and unable to vindicate the *categorical* applicability of recognisably moral rational requirements. Again, we don't have a vindication of normativity in the rationality-involving sense that is recognisably moral. Error theory is still on the cards.

Unfortunately, error theory is still not off the hook. For there is one final variation on Smith's Entailment Argument. Call it the *Entailment Argument from the Goodness of Rational Requirements*.

Why follow rational requirements rather than just voices in one's head? The answer seems to be: because there is something *good* about them. If there wasn't anything good at all about following rational requirements rather than voices in one's head then the argument which I make in this chapter (viz., that agents with the capacity for rational deliberation should follow the rational requirements like IP and Reliance and not voices in their heads) breaks down. For then it would be just as permissible to follow voices in one's head rather than rational requirements. But I can't accept that this argument breaks down, so I should argue that there is something good about rational requirements. With that accepted a further claim to be made is that the goodness of rational requirements cannot be relative to times or agents. For it would result in a very strange view if we admitted that it would only be good-for-Abe to abide by IP but that it need not be good-for-Bert to do that, given that Abe and Bert have relevantly similar psychologies.<sup>439</sup>

---

<sup>439</sup> Markovits (2011: 264n23).

So far no problem for error theory. But now look at Abe considering whether he should kill Bert. It would be inconsistent for him to continue to take the means to *his* ends out of recognition of the fact that it is good to follow the Instrumental Principle *full stop* and nevertheless to perform the action of killing Bert that results in a state of affairs in which something that is goodness-producing—Bert and his ability to follow IP—gets annihilated. Either following IP is objectively or non-relatively good, in which case agents that can follow IP should be respected, regardless of whether these agents are identical to oneself or not. Or it is not universally good, but that leads to the weird view that it might be good-for-Abe to follow IP but that it need not also be good-for-Bert to follow IP.

Both horns of the dilemma are unacceptable for my error theory. The second horn, I argued two paragraphs up, sits ill with my defence of Proceduralism+, which requires that there is something objectively good about the rational requirements that it postulates. But that leaves us with the first horn, which, unfortunately for error theory, entails that agents should respect other agents as a means to respecting the objective goodness of simple rational requirements like IP. And this ‘should’ is clearly recognisable as a moral should.

To this way of making the argument the error theorist can respond as follows. She can insist that talk of the *goodness* of following rational requirements is in fact too much out of line with the proper functioning account of rational requirements. Recall from §6.2.1 the distinction between the proper functioning account of rational requirements and the *intrinsically valuable status account* of rational requirements. According to the latter, what grounds the truth of for instance the Instrumental Principle is that failing to obey a rational requirement involves somehow undermining or failing to *respect the valuable status of being rational*; i.e., this would involve failing to respect something that is *good*. But the proper functioning account doesn’t mention the intrinsic value or goodness of obeying rational requirements. It only uses the idea that obeying rational requirements is the only way to perform one’s function as a being with a psychology like ours. So the error theorist relying on Proceduralism+ can question the legitimacy of this talk of the goodness as it is employed in the current objection. This means that the objection just discussed that relies on it can be rejected as well.

To sum up, there are no rational requirements with recognisable moral content that are entailed by the rational requirements that Proceduralism+ vindicates.

#### **6.4.2 Can Categorical Moral Norms be Grounded in the functional Profile of Human Psychologies?**

In this sub-section I look at the objection that my description of the functional profile of our psychologies, and the attitudes of desiring and believing that comprise it, is somehow mistaken. In essence the objection is that a different account of the functional profile of human psychologies provides the building blocks for a vindictory explanation of moral normativity in the rationality-involving sense.

Above I argued that desires are the kinds of mental states that ‘aim at their own satisfaction’ or ‘are dispositions to realize their own content’. I also argued that this interpretation of what desires are gets us an argument for error theory, as the only rational requirements that can be grounded in that description are those that do not have recognisable moral content.

But there are different descriptions of what desires are. Perhaps desires are the kinds of things that operate *in a law-like way* and are *in virtue of that* subject to *universalization constraints*.<sup>440</sup> There are various ways of spelling out this idea, travelling from Kant through Hare to contemporary philosophers like Korsgaard and Velleman. The main idea is that agents who have a desire to kill or steal will have desires that they cannot consistently will to have in virtue of the following rational requirement:

<b>Universal</b>	RR (if you have and exercise the capacity for practical rationality, then the only desires you have are such that they can be fulfilled in a way consistent with the fulfilment of all other desires—including desires of other agents—or that are such that their content is suitably universal)
------------------	---

Universal has two elements. I discuss them in turn. Agents are not allowed to hold a desire if it turns out that this desire frustrates the satisfaction of other desires, regardless of whether these are instantiated in them or in other agents. So for instance, Celine cannot kill Derek because although Celine’s killing Derek would satisfy one of her deep desires, that deep desire is also inconsistent with the idea that

---

<sup>440</sup> Smith (2010: 127, 2012a: 240); Markovits (2011: 277); Copp (2007: Ch. 8).

the fulfilment of no desire may frustrate that of other desires, *wherever instantiated*. And the satisfaction of Celine's desire to have Derek killed frustrates those desires that are instantiated in Derek—indeed, all of them. That was the first aspect of Universal. The other aspect of Universal is that desires can also be said to be suitably universal when their content is universal. That is, agents can desire equal income for all rational agents alike; but they can't desire, for instance, income discrepancies based on non-rational features of agents such as eye-colour.

Various philosophers add different emphases to this picture. Korsgaard for instance thinks that in performing the action of Killing Derek, Celine would not be able to recognize herself *as an agent*. The reason why desires operate in law-like ways and are subject to universalization constraints is that only that allows agents to recognize themselves as agents. Velleman adds to Universal the idea that it is constitutive of actions that they are intelligible, and that the only way in which this aim can be realized is when Universal is respected.<sup>441</sup>

The problem with the Universalization strategy however is that the thicker account of the functional profile of psychologies that it necessitates can't be accepted. The rational requirements that Proceduralism+ postulates are inescapable because they are tied to what beliefs and desires *are*, and these are mental states that we *cannot help to have*. But the rational requirements like Universal that Korsgaard and Velleman postulate are only *meant to be* inescapable because we need to be able to recognize ourselves as agents, or because our actions must be intelligible. However, in fact they are escapable. Although I simply cannot get rid of all my desires and beliefs, I *can* get rid of the idea that I need to understand myself or that my actions have to be able to be intelligible. So recall from §5.3.2 my distinction between *legitimate* and *illegitimate* 'so what?'-responses. Legitimate 'so what?'-responses are 'so what?'-responses that make sense (cf. Joyce's example of the unwilling gladiator). Illegitimate 'so what?'-responses don't make sense, and I argued that examples of such 'so what?'-responses are given by rational requirements like the Instrumental Principle and Belief Consistency. After all, it doesn't make sense to shrug off requirements on the formation of mental states that one cannot help but have whereas it does make sense to shrug off requirements on

---

<sup>441</sup> Korsgaard (1996), Velleman (1996, 2000, 2004). Also see Copp (2007: 276); Silverstein (2012: 8).

actions, such as refraining to throw sand, that are entirely arbitrarily and unwillingly imposed on oneself. The question is which way Universal goes, and my claim is that Universal is like the rule of gladiatorial combat, not like Instrumental Principle and Belief Consistency. After all, it makes sense to shrug off the demand for understanding oneself—it is not the case that one cannot but want to understand oneself—whereas it doesn't make sense to shrug off the demands of IP *et cetera* because it *is* the case that one cannot but have desires and beliefs. So a thicker notion of the proper function of the human psychologies isn't defensible. And so rational requirements like Universal, which have recognisable moral content, can't be grounded in it.<sup>442</sup>

This is a versatile reply. Exactly the same answer applies, *mutatis mutandis*, to suggestions other than that of Velleman and Korsgaard. Take my suggestion from §6.2.2 that it might be part of the functional profile of our psychologies that our psychologies are the kinds of things whose function it is to allow us to get along with each other. Again, the question is whether this is an 'aim' or 'function' that it makes sense to shrug off. And the answer is that it is. We all have desires and beliefs and it doesn't make sense to shrug off the requirements that govern them. But some of us are at least some of the times powerful enough, or stealthily enough, to perform actions that undermine (mutual) trust in society. For those people, I surmise, it makes sense to shrug off the alleged requirement on our psychologies that it must produce actions that are such that they, precisely, do not have the effect of undermining trust in society. They can get more of what they want without this having any negative effects on them.

One possible reply to this line of reasoning is that I have so far focussed exclusively on different understandings of notions of *agency*, not on the different accounts of the function of individual desires *per se*. According to the Universalization account, desires only *count* as desires when they are subject to a universalization constraint. The advocate of a success theory might insist that *that* idea has so far not been undermined. All we have undermined are Korsgaard's and Velleman's *additions* to the bare idea of Universal guiding the formation of desires; viz., the ideas about recognizing ourselves as agents and the intelligibility of action. The reply is that this doesn't disprove that *desires* simply are the kinds of things that

---

<sup>442</sup> Also see Silverstein (2012: 5).

are subject to rational requirements like Universal.

This is correct as far as it goes, but it foregoes two important points that speak in favour of the error theorist. First, we need an *explanation* of why desires are such that they must operate in a law-like way. Compare this to my own claim that desires are the kinds of things that aim at their own satisfaction. My claim is very minimal and looks platitudinously true. It sounds very strange to assert that one can have a desire for something in such a way that the satisfaction of that desire isn't part of one's state of mind. But the 'desires must operate in a law-like way'-idea goes further than that—it goes beyond this platitudinously true claim—and we can rightfully ask for an explanation of this further claim. That is the first point. The second point is that Korsgaard and Velleman provide the required explanations: desires are subject to rational requirements like Universal because if they weren't we wouldn't be able to understand ourselves as agents or else our actions wouldn't be intelligible. But these explanations were unsuccessful. And we have also already seen that my objection to Korsgaard's and Velleman's explanations is very versatile, and so it is plausible to believe that *other* explanations of why it must be part of desires that they operate in a law-like way will also fail. Therefore, merely looking at desires *themselves* in an attempt to argue that they are the kinds of things that are subject to universalization constraints won't work.

To conclude, therefore, the friend of a success theory of morality won't find solace in looking at a different, and richer, account of either our psychologies or the various mental states that comprise them.

## 6.5 Looking Elsewhere

How about grounding categorical moral normativity in the rationality-involving sense not in the functional profile of mental states and psychologies, but in something else entirely? Perhaps we can say that what supplies morality's categorical moral normativity in the rationality-involving sense are *reasons* that are 'part of the fabric of the world' in much the way in which Mackie believed that morality's 'objective prescriptivity' would have to be 'part of the fabric of the world'. However, in this section I argue that such attempts fail.

Start by distinguishing between robust and quietist non-naturalism.<sup>443</sup> Robust non-naturalists argue that there exist irreducibly normative and non-natural properties in the same way that natural properties like being a £20 note exist.<sup>444</sup> Quietist non-naturalist realists also argue that there exist irreducibly normative properties, but they deny that they exist in the same way that natural properties like being a £20 note exist. Moral properties are not real but *irreal*. This is meant to guarantee moral properties' metaphysical innocence.<sup>445</sup> I will first argue that robust non-naturalism fails. After that I argue that irrealist theories face the same conclusion.

In arguing against non-naturalism I will consider Enoch's recent views, and then extrapolate from this discussion.

Enoch offers the following argument for robust ethical non-naturalism. Whether we can accept robust ethical realism "ultimately depends on ... [whether] its plausibility score is on the whole higher than that of any competing view."<sup>446</sup> What this means is that if robust ethical realism can make the best sense of our platitudinously true claims about ethics and if it has various arguments supporting it without taking too many blows in the process—being metaphysically extravagant, being in fact unable to explain our epistemic and semantic access to moral properties, *et cetera*—then it will, overall, be preferable to competing theories about ethics, including the error theory and various naturalist and expressivist theories. And, the argument continues, the antecedent of this conditional is true, so robust ethical realism should be accepted.

As Enoch also suggests, authors of philosophical treatises can, and should, be forgiven for not going through the actual exercise of comparing plausibility points (and, in any case, this would require a fairly settled method for tallying plausibility points that we currently don't have).<sup>447</sup> So what I am going to do instead, in the spirit of Enoch's own method of argumentation, is to supply some

---

<sup>443</sup> Olson (forthcoming-a). I look at non-naturalism because I have already argued against naturalist views in the previous chapter.

<sup>444</sup> Enoch (2011a).

<sup>445</sup> Skorupski (2010); Parfit (2012); Scanlon (2009). Also see Kornblith (2012: 119).

<sup>446</sup> Enoch (2011a: 267).

<sup>447</sup> Enoch (2011a: 267, 14-5).

considerations that count in favour of accepting error theory rather than robust ethical realism. This discussion is meant to show that robust ethical realism suffers from more problems than one might think. I also explain what the plausibility points are that error theory has, and the combined result of this is that we should conclude that error theory is more plausible than non-naturalism.

The first thing I want to mention is that both robust realism and my error theory save the appearances, at least to a very large extent. True, robust realism saves *more* appearances than error theory—it saves *moral* appearances—but other than that, I have argued, error theory saves all the other appearances, including epistemic and prudential ones. So the fact that error theory only doesn't save the *moral* appearances is, albeit something that counts against it, not something that counts too heavily against it. Indeed, I surmise that this loss in plausibility points may be roughly on par with the loss in plausibility points that robust realists face with their explicit acceptance of a non-naturalist metaphysics. With Enoch, then, I believe that error theory and robust realism are each other's most formidable opponents.<sup>448</sup>

Second, on what I think of as a crucial issue—the *explanation* of the normativity of various facts, be they moral, prudential, epistemic or what have you—my account is to be preferred over Enoch's. Enoch agrees with me that moral normativity would have to be of the rationality-involving kind and would have to have categorical force to *count* as moral normativity.<sup>449</sup> He also agrees with me that a purely naturalist metaphysics can't underwrite or ground this kind of normativity.<sup>450</sup> Moreover, Enoch, like me, has a vindicatory explanation of the normativity of those facts he believes are normative and exist. I have presented my own proper functioning account of rational requirements; Enoch presents his argument from deliberative indispensability to 'Robust Metanormative Realism' in Chapter 3 of his *Taking Morality Seriously*. But as I will now argue, although I believe that my own account is successful, Enoch's is not.

---

<sup>448</sup> Enoch (2011a: 115).

<sup>449</sup> Enoch (2011a: 94).

<sup>450</sup> One caveat: Enoch is happy to be called a naturalist on the following, different understanding of naturalism: if you accept, as Enoch does, "the (perhaps strong) supervenience of the normative on the natural" then you are a naturalist (2011a: 101n2).



I can accept Enoch's "Argument from the Deliberative Indispensability of Irreducibly Normative Truths" to an extent.<sup>451</sup> That is, I can accept that we need to postulate irreducibly normative truths in order to make sense of our practical deliberation. This, indeed, is similar to the kind of argument I gave in the previous chapter for why we have to conceive of the normativity of morality in the rationality-involving sense (and to have categorical force). And given that robust realism and my error theory save a large number of the appearances, a lot is going to turn on the metaphysical respectability of Enoch's view.

Here is where we encounter problems. In his chapter on metaphysics Enoch rejects what he calls the Sheer Queerness worry; viz., the worry that we can infer from the (alleged) sheer queerness of moral properties that they don't exist.<sup>452</sup> Enoch rebuts this worry by insisting that the world is a queer place—Enoch asks us to consider, with Platts, neutrinos, aardvarks, infinite sequences of objects and impressionist paintings—and concludes that since we don't put nuclear physics, zoology, formal semantics and art history into disrespect in light of the queerness of the things that these disciplines talk about, we shouldn't do that with morality and its queer properties either.<sup>453</sup> I don't think that this inference holds. The question we should ask, in line with my proposal for thinking about naturalness in §6.2.2, is whether the properties of being an aardvark, an expressionist painting *et cetera* stray too much from archetype natural properties. And the answer is that none of these properties do. But the *normative* magnetism of moral properties, given the absence of a story about the function of a psychology that they can be grounded in (recall the result from §6.4.1-2), does.<sup>454</sup> That is, although we can understand what it is for something to be normative when we have at our disposal a story about the function of a psychology that grounds the truth of the normative claims involved and although we can see that this story is sufficiently similar to a story about other properties with functions (such as being a £20 note) that are clearly naturalistically respectable, it is very hard to understand how a property that is normative in and out of itself can be naturalistically respectable.

---

<sup>451</sup> Enoch (2011a: 50). See Enoch (2011a: 116) for a statement of the claim that error theorists can accept this argument.

<sup>452</sup> Enoch (2011a: 134-6).

<sup>453</sup> Enoch (2011a: 135); Platts (1980: 72).

<sup>454</sup> Also see Scanlon (2009 Lecture 2: 30); Gibbard (2003: 32).

This is queerness understood as ‘mysteriousness’ (see §2.2.3). In that section I argued that Mackie uses a different understanding of queerness (being *unusually unusual*) and that a quick glance at understanding being queer as being mysterious doesn’t seem to get us a working argument of queerness either. But in fact queerness *can* be understood as mysteriousness, as follows. What is important in the context of many theories, both scientific and philosophical, is that we can provide a vindictory explanation of the existence of a theory’s postulates, and, as I argue in this chapter, although we can provide a vindictory explanation of the existence of the norms of Proceduralism+ (because these can be grounded in a functional story of human psychologies and its constituent parts), we can’t provide a similarly successful vindictory explanation of the existence of mind-independent moral properties that are nonetheless normative and indeed able to affect agents’ (practical) rational deliberation. Now, postulating such properties anyway would get us a success theory of morality, but only at the cost of, precisely, a mysterious (because explanatorily vacuous) account of the nature of those properties. For in the same spirit we could explain the existence of witches by postulating the existence of the property of being able to curse innocent children so that they develop diseases *because* one has cursed them. But in the absence of an explanation of what it is in virtue of which things can have that property, this ‘explanation’ of the existence of witches clearly doesn’t licence a success theory of witch discourse. In essence, my objection to moral realism of the robust, non-naturalist variant goes like this.

How does this relate to Shepski’s complaints that we can’t infer from our failure to understand something that it doesn’t exist (cf. §2.2.3)? Shepski’s complaint doesn’t carry a lot of weight as he merely asserts that this inference is unwarranted (and that an inference from a failure to be able to conceive of something might work, but that this isn’t a queerness worry).<sup>455</sup> I disagree. Properties for which we cannot provide a vindictory explanation but which we postulate anyway to vindicate the existence of some other phenomenon (witchcraft, morality) are clearly and in a good sense ‘queer’.

So this response of Enoch’s to the queerness of moral properties doesn’t work. And again given the plausibility of my error theory when it comes to providing vindictory explanations of normativity elsewhere, it seems to me that

---

<sup>455</sup> Shepski (2008).

robust ethical realism should be rejected in favour of my error theory. A view that isn't queer and only fails to vindicate a small portion of our normative talk—viz., our moral talk—is to be preferred over a view that vindicates that extra small portion of our normative talk, but is queer.

Enoch has a reply to this, however. He claims that his arguments for the existence of normative (ethical) truths “were *not* ... intended merely as attempts at fleshing out the commitments of moral and normative discourse. They were meant as attempts to *establish* Robust Realism.”<sup>456</sup> Since (deliberative) indispensability arguments can have ontological implications, the Argument from Deliberative Indispensability simply shows that irreducibly normative truths (facts or properties) exist. In response, the error theorist should insist that indispensability arguments work *so long as* their ontological implications do not flout a general Metaphysical Naturalism, but that, unfortunately, as I've just argued, Enoch's Argument from Deliberative Indispensability flouts that very thesis.

To this move Enoch has a further reply. Enoch puts forward the Moorean Objection against error theory, according to which the fact that one is comparatively more certain of a certain existential claim compared to a sceptical claim justifies one in retaining the existential claim.<sup>457</sup> The objection is inspired by Moore's tactic to argue against the skeptic about the existence of an external world:

I cannot help answering [sceptical doubts about the existence of an external world as follows]: It seems to me *more* certain that I *do* know that this is a pencil and that you are conscious, than that any single one of these [sceptical] assumptions is true, let alone all [of those assumptions]<sup>458</sup>

In the moral case we would have that it is more certain that one knows that killing is wrong than that it isn't because the sceptical 'assumption' that is error theory is true.

The Moorean objection faces the following dilemma.<sup>459</sup> Either Moore's claim is a psychological claim or it is a normative claim. If it is a psychological claim—as a matter of fact, people continue to believe in moral realism even when presented with arguments for error theory—then it seems to have no probative force. Philosophical and intellectual inquiries are concerned with what we *ought* to believe

---

<sup>456</sup> Enoch (2011a: 116).

<sup>457</sup> Enoch (2011a: 117-21).

<sup>458</sup> Moore (1953: 143).

<sup>459</sup> Daly and Liggins (2010: 225).

and given that aim it doesn't matter what some or even all people believe or what they are able or unable to believe. If on the other hand the Moorean objection presents a normative claim then we need reasons to accept it. Suppose, then, that common sense claims are "more certainly true than any evidence that is brought against them."<sup>460</sup> However, the problem here is that even if this is true then this presumably isn't a brute fact but that the Moorean Objection fails to provide the required explanation. So Moore's argument fails.

Enoch though embraces the second horn of this dilemma. He therefore accepts that Moore's claim is normative. But Enoch insists that Moore's argument isn't explanatorily vacuous or question-begging, for Enoch argues that there *must be* easy knowledge; i.e., knowledge of the form:

This tomato appears red to me (by introspection); This tomato is red (by perception); So my color-perception got it right this time; (Similarly for other things); Therefore, my color perception is reliable<sup>461</sup>

Otherwise, argues Enoch, skepticism about the external world can't be avoided. But, to me it seems that this requires further argumentation. There are many responses to external world skepticism out there that do not rely on easy knowledge.<sup>462</sup> I think we are therefore back to a consequence of easy knowledge that Enoch himself recognizes: Bootstrapping. Appeals to easy knowledge are unjustified, and so this final attempt to undermine error theory fails.

As far as the debate stands at this point it seems to me that error theory emerges victorious: it is better to have a fully naturalistically respectable view that condemns a small part of our normative talk—moral talk—to systematic error than to have a view that is metaphysically queer but saves moral talk. It would be better if I had a knock-down argument against non-naturalism as such though, but I don't. It would also be better if I had the space to compare my error theory with all the other version of non-naturalism on offer, but I don't have that either. Finally it would be better if we had a better sense of how to tally plausibility points. Nevertheless,

---

<sup>460</sup> Armstrong (2006: 160).

<sup>461</sup> Enoch (2011a: 119).

<sup>462</sup> Greco (2007).

similar kinds of considerations will apply to other non-naturalist theories. Hence I conclude, at least provisionally, that error theory fares better than non-naturalism.

I turn to irrealist theories. I discuss, in turn, the theories of Scanlon and Parfit. After this I generalize my findings to other theories.

*Scanlon.* Scanlon accepts that what is normative in the practical realm always involves reasons. He distinguishes, as is customary, between the concept and the property of a reason and his view does not reductively analyse reasons on either level. So Scanlon's non-reductionist view does not say that what we really mean when we say 'X has a reason to  $\phi$ ' is that X stands in a purely non-normative relation to purely non-normative facts. It also doesn't say that the *property* of a reason can be so analysed, for instance by saying that reasons are (the 'are' of identity) purely non-normative facts or relations.

Off the cuff, these commitments seem to imply a robust non-naturalist metaphysic. But one way to slim down the metaphysical commitments of this theory whilst staying non-reductionist about moral normativity in the rationality-involving sense on both the conceptual and the property level is to deny that the property of a reason exists *in the world*. This is Scanlon's proposal. Moral reasons, albeit intrinsically normative, "constitute a distinct realm and [yet] need no metaphysical reality."<sup>463</sup>

Why don't moral reasons require a distinct metaphysical reality? Start with some very general reflections about ontology. With Quine, Scanlon argues that the way to understand the ontological commitments of a set of statements is to first translate these statements into the language of first-order logic, and then to determine what things there must be in the "universe of discourse" of a model in which all of these statements are true.<sup>464</sup> These things are what we are ontologically committed to in accepting those statements. Importantly, the 'universe of discourse' isn't itself anything ontological but "merely formal: a way of representing all those things that are presupposed by some set of statements, about the natural world or

---

<sup>463</sup> Scanlon (2009, Lecture 2, p16); also see Parfit (2012 Vol2: 479, 486-7); Skorupski (1999: 438).

<sup>464</sup> Quine (1948). Also see Burge (1992: 17).

about anything else.”<sup>465</sup> So the universe of discourse is not the universe itself; that is, it is not the world around us as studied by the physical sciences. This distinction allows Scanlon to argue that the ontological commitments of a first-order discipline such as mathematics and morality only have implications for one’s ‘universe of discourse’ and not for ‘the universe’. First-order disciplines that do have implications for the universe are limited to the natural sciences and other disciplines, such as those studying witches and spirits, that can only be true, if they are, if they entail commitments about the universe. But this doesn’t hold for (moral) reasons. These can be true without entailing commitments about the inventory of the universe.

In making this argument Scanlon is committed to a permissive as opposed to a restrictive first-order view of ontological commitment. His permissive view says that we should decide what sentences to accept as true by applying the criteria appropriate to the relevant first-order disciplines (practical and moral thinking included) and then accept as existing whatever one is quantifying over in one’s ‘universe of discourse’. According to such more restricted views “we should ... avoid ontological commitment to anything other than physical objects”.<sup>466</sup> And because it seems that normative truths don’t fit into *the universe*, such restricted views are incompatible with normative truths about what we have reason to do.

Clearly to get a success rather than an error theory we should accept the permissive view. But what justifies accepting the permissive view? Scanlon’s answer is that doing this allows us to countenance whole rafts of claims as true (viz., claims about mathematics, claims about normativity, *et cetera*) and because in doing so we have nothing to lose. In particular, we are not losing our allegiance to metaphysical naturalism.

I argue that this is all well and good, but that some questions remain. One question is; where, exactly, does the *normativity* of reasons come from on this picture? According to Scanlon, the “distinctive aspect of normative truths is ... a matter of what Quine called “ideology” (the predicates we employ) rather than

---

<sup>465</sup> Scanlon (2009: Lecture 2, p4).

<sup>466</sup> Scanlon (2009, Lecture 2, p2).

ontology (the things we quantify over).”<sup>467</sup> This suggests that the normativity of reasons comes from our predicates, not from the referents of these predicates. However, this is puzzling. For it obfuscates the need for an elaborate irrealist theory of the referents of reason predicates.

So perhaps this is not what Scanlon has in mind. Perhaps then we should say that what is normative are our reasons, not our predicates that we use to talk about our reasons. Nevertheless, this then invites the question: how do these reasons exert normative force *over us*? The problem is that we can imagine *counter-reasons*; ‘reasons’ that exist in the ontologically light-weight sense of existence that Scanlon favours, and which we can imagine to be very similar to that of reasoning albeit resulting in different conclusions. After all, if we can justify the existence of reasons by reference to the internal practice of reasoning alone—and thus without taking considerations about what forms part of the *universe* into account—then we can justify the existence of counter-reasons by reference to the internal practice of counter-reasoning alone.<sup>468</sup> So suppose that we end up with a reason not to kill and a counter-reason *to kill*. Given that both reasons and counter-reasons exist, we can ask: why should we heed to our *reasons*? How do reasons exert their normative force over us given that there are also counter-reasons that rogues like Jack the Ripper could appeal to? Scanlon has no principled answer to this, and indeed he can’t have such an answer as reasons and counter-reasons are equally respectable given that all it takes to have them is some practice of thinking about them.<sup>469</sup> Now Scanlon could try to get the required difference between reasons and counter-reasons by inflating his account of existence to a more “platonistically friendly” one.<sup>470</sup> But that would entail a non-naturalist view inferior to error theory as I have already argued earlier on in this sub-section. Overall, then, Scanlon’s irrealism doesn’t work. I will now argue that Parfit’s version of this theory faces the same conclusion.

*Parfit.* Parfit holds a view similar to Scanlon’s by arguing for a view he calls Non-Metaphysical Non-Naturalist Normative Cognitivism, according to which

---

<sup>467</sup> Scanlon (2009, Lecture 2, p13-4).

<sup>468</sup> Enoch (2011a : 125).

<sup>469</sup> Enoch (2011a : 126).

<sup>470</sup> Enoch (2011a : 124).

“there are some claims [about reasons] that are, in the strongest sense, true, but these truths have no positive ontological implications.”<sup>471</sup> Like Scanlon, Parfit starts by looking at what kind of view about existence in general he can or should accept. He accepts Possibilism, according to which some things are not real but nevertheless exist. Parfit accepts Possibilism on independent grounds, but also because Possibilism can get combined with the Plural Senses View of Existence, according to which there are different senses in which things can exist. This package of claims allows Parfit to claim that although simple Metaphysical Cognitivism is true for some areas of discourse (discourse about rocks, which entails robust truths about rocks that require us to answer questions about what exists in the ontological sense of ‘exist’), there are other areas of discourse, such as those about reasons for action, which involve robust truths that do not require us to answer questions about what exists in the ontological sense of ‘exists’.<sup>472</sup>

Again the same issue of the existence of counter-reasons applies. For again we can consistently with Parfit’s story come with a story about the (non-ontological) existence of counter-reasons that problematizes his view. After all, which one has real authority—reasons or counter-reasons; i.e., the reasons that we accept not to kill or the counter-reasons that Jack accepts *to steal*? Parfit can’t argue that it is reasons that we ought to accept, unless, perhaps, he accepts that these exist in a more ontologically laden way. But that brings out the problems for non-naturalist theories.

Both theories suffer from the same objection, and in virtue of the same feature of their theories; viz., that they refuse to give their theories robust metaphysical import. It seems safe to assume that other quietist theories, such as for instance Dworkin’s, suffer from the same problem.<sup>473</sup> Without going through all of these different theories though I conclude, provisionally, that quietist realism can’t work.

Neither non-naturalist nor quietist realism can save the day for the moral success theorist in the present context in which Proceduralism+ and its rejection of

---

<sup>471</sup> Parfit (2012 Vol2: 479).

<sup>472</sup> Parfit (2012 Vol2: 487). Metaphysical Cognitivism is the view that there are claims that are, in the strongest sense, true and where these truths have positive ontological implications.

<sup>473</sup> Dworkin (1966).



rational requirements with categorical moral force is taken as accepted. At this point, therefore—and remembering our rejection of non-cognitivism in Chapter 1 and that of naturalism in Chapter 4—moral error theory has become the most plausible metaethical theory. No moral statements attributing moral properties are true.

## **6.6 Conclusion and Preview**

This chapter has argued that the kind of categorical moral normativity that our ordinary moral discourse non-negotiably commits us to does not exist, at least not in the actual world. In conjunction with the result of the previous chapter that we can only have morality when this kind of normativity exists, a moral error theory has been established, at least for the actual world. Importantly, this moral error theory, I argued, has a convincing reply to the problem that I called the Formulation Dilemma. It has better arguments for the Non-Negotiable Commitment Claim and the Substantive Claim than existing error theories like that of Mackie and Joyce. The next chapter discusses objections to error theory that I haven't engaged with so far. I argue that these objections don't hurt my error theory. The final chapter of this dissertation discusses the issue of how we (prudentially) should respond to the truth of my version of error theory.

## Chapter 7

# Further Objections to Moral Error Theory

### 7.1 Introduction

In the previous chapters I argued that moral error theory is plausible. It is coherent and can deal with various objections. This chapter formulates and rejects objections to error theory that I haven't discussed before. This amounts to further support for my error theory.

One objection I haven't dealt with yet is that since we can't honestly believe error theory in our least philosophical and most commonsensical moments we should abandon error theory (§7.3). Another such objection is that my error theory of categorical moral normativity in the rationality-involving sense isn't *pervasive* enough. Perhaps there are no moral right and wrongs, but if moral goodness, values and other aspects of morality are not affected by my arguments then we can doubt whether error theorists can rightly say that moral discourse 'as such' is in error (§7.4). A further objection is that we can infer from (i) the existence of a possible world identical to the actual world in all non-moral respects in which there are moral properties and (ii) a suitably formulated non-arbitrary constraint on moral properties' instantiation conditions that (iii) the actual world also contains moral properties, and hence that error theory is false (§7.5). I close with a summary and preview (§7.6). But I start, in the next section, with two objections that require just a couple of paragraphs to deal with.

### 7.2 Quick Objections, Quick Replies

Consider the principle of epistemic conservatism:

a person is to some degree justified in retaining a given belief just because that person has that belief<sup>474</sup>

---

<sup>474</sup> Daly and Liggins (2010: 223).

On the basis of this quote we can formulate the objection from epistemic conservatism. Consistent with the quote we can say that the degree of justification for a first-order moral claim such as Wrong ('stealing is morally wrong'—see §3.2), which we possess just because we have the belief that Wrong is true, is greater than the degree of justification we have in the truth of error theory. From this it follows (at least on the assumption that we should believe what we have the greatest degree of justification to believe) that we shouldn't believe error theory.

Error theorists can either claim that the principle of epistemic conservatism carries no weight at all or that it carries some weight.<sup>475</sup> The former, hard-line response is possible, but difficult to defend, and error theorists don't need to defend it. For instead they can accept that the principle carries some weight but insist that the balance of evidence favours error theory nonetheless. This thesis as a whole aims to establish this. The reader must be the judge.

The second objection we can dismiss rather quickly starts from the observation that when we do philosophy we are trying to reach a reflective equilibrium between our pre-theoretical beliefs about various things on the one hand and the (philosophical) theories we construct to explain these things on the other hand. The objection is that error theory can never feature in such a reflective equilibrium because it denies that some pre-theoretical beliefs, such as the belief that stealing is wrong, are true.<sup>476</sup>

This objection fails as well. It asserts on unwarranted *a priori* grounds that no reflective equilibrium that contains error theory is forthcoming. This remains to be seen and this dissertation aims to show that error theory merits the kind of epistemic warrant that justifies overthrowing beliefs like Wrong.

In light of this advocates of this objection may shift ground and assert instead that although error theory might achieve reflective equilibrium, there is unlikely to be a unique reflective equilibrium. Moreover, they can say, if one of these equilibriums doesn't contain the error theory then that licences embracing that particular equilibrium simply because in it error theory is false.

This objection also fails. We need an *additional* argument for why the set of claims that doesn't include error theory and yet reaches reflective equilibrium is to

---

<sup>475</sup> Huemer (2005: 99); Daly and Liggins (2010: 223).

<sup>476</sup> Daly and Liggins (2010: 216).

be embraced rather than another set of claims that also reaches reflective equilibrium but does include error theory. That argument needs to go beyond considerations of mere reflective equilibrium. Therefore, the objection from reflective equilibrium must collapse in one of the other arguments against error theory that are discussed in this chapter or elsewhere in the dissertation. But these arguments are not successful, as I will now continue to argue.

### 7.3 Honesty

Consider the following maxim of honesty:

never put forward a ... theory ... you cannot yourself believe in your least philosophical and most commonsensical moments<sup>477</sup>

On the basis of this maxim we can formulate the Objection from Honesty. On the assumption that we cannot believe error theory in our least philosophical and most commonsensical moments we shouldn't put forward (or we have sufficient reason to reject) error theory.

Daly and Liggins argue that this objection doesn't work.<sup>478</sup> First (1) we should not underestimate what we can bring ourselves to believe. Indeed, unless there is countervailing evidence, we should take those philosophers who claim that they believe moral error theory and other 'unbelievable' claims (such as the compositional nihilist's claim that there are no chairs) at their word.<sup>479</sup> If (1) is true then we *can* believe error theory, in which case the maxim of honesty no longer threatens the error theorist's position. Second (2) the maxim misses its target because in philosophy our aim is to determine what we ought to believe and it is

unclear why our capacities for forming beliefs should constrain our thinking about what we ought to believe<sup>480</sup>

So if (2) is true then the objection from honesty fails because it makes a claim about what we in fact cannot believe whereas what is relevant in the debate about error theory is a claim about what we *should* believe.

---

<sup>477</sup> Lewis (1986: 135).

<sup>478</sup> Daly and Liggins (2010: 216).

<sup>479</sup> See Merricks (2001) on compositional nihilism; see Pigden (2010: 17) on moral nihilism.

<sup>480</sup> Daly and Liggins (2010: 216).

As it stands, this two-fold reply fails to convince. First, reply (2) fails to appreciate that the maxim of honesty can also be interpreted as a normative claim, as follows: you ought to believe only what you can bring yourself to believe. This normative claim may be false, but it is a normative claim nonetheless, and it gets us a clash between two normative principles; viz., the maxim of honesty and the maxim according to which we should believe what we have most evidence to believe. But if that is so then the question becomes: which of these normative principles should we accept? And the worry for error theory is that if the first principle carries more normative weight than the second then it follows that we should reject error theory.

Worse, as stated Daly and Liggins' first reply fails to convince as well. For Daly and Liggins do nothing to make plausible the claim that when people believe error theory, they are at least some of the times *in their least philosophical moments* (circumstances C) and not always in circumstances C' (their philosophical moments). The problem is that for all we know, the set of people that are error theorists in circumstances C' and are able to hold on to the belief that error theory is true in circumstances C is empty. And this may be so *even though* it is true that (1) we should not underestimate what we can bring ourselves to believe and indeed that some people do in fact believe the error theory, whom we should take at their word unless we can cite, as Pigden puts it, "brain injury or massive self-deception."<sup>481</sup> So reply (1), because it fails to specify the conditions under which error theory is believed, cannot be used to resist the objection from the maxim of honesty.

Fortunately Daly and Liggins' response can be strengthened. To strengthen their first response, I can note that, unless I suffer from brain injuries or other delusions (for which there is no evidence), I am the best authority on what I believe and in what circumstances, and I can assure you that I believe my version of the error theory both in circumstances C' and in circumstances C. So the abovementioned set of believers in error theory is non-empty. Having thus strengthened this response to Lewis' honesty maxim I have at the same obviated the need to strengthen the other response. For even if it turns out that Lewis' maxim carries more weight than the maxim of belief, if we can believe the error theory in both C and C' then the maxim of honesty simply doesn't apply to error theory.

---

<sup>481</sup> Pigden (2010: 17).

## 7.4 Pervasiveness

The second larger objection I discuss is that my error theory fails because it debunks, at most, the deontic features of moral discourse. Perhaps moral statements that contain concepts like WRONG (equivalent to ‘ought not’) and RIGHT (equivalent to ‘ought’) are in error. But this leaves statements containing evaluative concepts like GOOD as well as thick concepts like RUDE error-free. And the same might hold for imperatives. After all, my argument in Chapter 5 only showed that moral statements such as Wrong metaphysically entail the moral reality thesis —nothing at all has been said about:

Good	Giving to famine relief is morally good
Rude	Scoffing at beggars is rude
Imperative	Down with war!

This gets us two questions. First, are these sorts of moral statements error-free? Second, if the answer to the first question is ‘yes’, does that endanger my conclusion that an error theory of moral discourse is warranted?

One option for error theorists is to admit that the answer to the first question is ‘yes’ but that this doesn’t entail that we should also answer the second question in the affirmative. We can hold the opinion that “the core of morality [identifies] ... some actions to be obligatory”<sup>482</sup> and claim that if the core of morality is error-ridden then error theory of moral discourse is licensed *even when* moral statements like Good, Bad, Rude and Imperative are not in error. A second option also answers ‘yes’ to the first question and ‘no’ to the second, but it does so without committing itself to the thesis that morality has a deontic core. According to the second option, it is already an interesting enough result that *a* part of morality is error-ridden; namely, its deontic part. A third option is to answer ‘no’ to the first question. We can argue that judgments like Good, Imperative and Rude are entailed by judgments like Wrong and that this means that these statements are in error.

I will argue that error theorists can embrace the third reply. This is good news. For the first reply opens up vexed issues about whether it makes sense to say

---

<sup>482</sup> Pauer-Studer (2009: 186).

that morality has a core and that its core is deontic.<sup>483</sup> And the second reply would concede too much to the success theorist: if for instance virtue-talk (cf. Rude) survives unscratched then virtue ethicists, with their claims about the *content* or *subject-matter* of moral talk and thought, are still in the running.

So how about the third option? Joyce writes that “talk of virtues ... generally implicates the existence of obligations”.<sup>484</sup> Virtues are often considered to be character traits that one is *obligated* to cultivate or that the virtuous agent is sensitive to when she acts in accordance with her moral obligations. But if that is so then virtue ethics requires deontological elements. And if accounts of the virtues leave out obligations-talk then they are too far removed from ordinary moral talk.<sup>485</sup> So in fact virtue ethics also paves the way for error theory.

How about Good, Rude and Imperative? Regarding Good, I borrow from Dreier and Joyce.<sup>486</sup> Morality consists of rules, argues the moral error theorist, but this might beg the question against evaluative conceptions of morality. For such conceptions of morality tells us what things are of value, not what the actions are that we should perform. However, some (if not most) evaluative conceptions of morality also tell us what to do after they’ve told us what is of value: they tell us to maximize value, or satisfice value, or what have you. This means that evaluative conceptions of morality often entail deontological conceptions of morality.

In fact, even those purely evaluative conceptions of morality entail deontological notions. For even if it is not obligatory to bring about morally good states of affairs, it would presumably still be good to bring it about, in which case one seems to have a *right* to perform the act that brings about the morally good state of affairs. And from that it follows that everyone else has a moral obligation to allow each other to perform that action.<sup>487</sup> The question, though, is whether there is anything that corresponds to *that* obligation, and if my error theory is right there isn’t.

---

<sup>483</sup> See, e.g., Scanlon (1982: 107); also see Schiffer (2003: 257-8); Lillehammer (2004: 762); Anscombe (1958); Bloomfield (2007); Joyce (2011b: 154); Macintyre (1981).

<sup>484</sup> Joyce (2001: 175).

<sup>485</sup> See Chapter 5 and Joyce (2001: 175).

<sup>486</sup> Dreier (1997: 81); Joyce (2001: 175-6). Also see Bond (1966).

<sup>487</sup> Hinckfuss (1987: 5).

This leaves us with thick moral concepts and imperatives. Consider thick moral concepts first.<sup>488</sup> Thick concepts somehow hold together descriptive and evaluative components, and everything we said thus far about purely thin evaluative and deontic concepts can be applied to the evaluative component of thick concepts. For instance, assume that the evaluative component of Rude is something like ‘undesirable but morally permissible’. If that is so then error theory directly kicks in concerning ‘morally permissible’. Moreover, it kicks in regarding ‘undesirable’ when we can link this evaluative conception of morality with the deontic one in much the way that Dreier and Joyce explained we can. Hence, utterances like Good and Rude carry links to deontological thin concepts. This is further evidence that error theorists needn’t fear pervasiveness worries.

Finally, let us look at imperatives. It can be argued that imperatives are merely used to express non-cognitive attitudes and that they are therefore not truth-apt propositions of the kind that error theorists balk at.<sup>489</sup> Consider ‘Down with war!’ Its grammatical form suggests that the person uttering it doesn’t *believe* that war is wrong and merely holds an unfavourable attitude toward war. The best reason for holding this view that I can think of is that if you *believe* that ‘war is morally wrong’ you will express that belief with a sentence in the indicative mood. For sentences in the indicative mood are standardly understood to express beliefs, and *in virtue of that* standardly understood to express willingness on the utterer’s part to take up certain commitments.<sup>490</sup> These commitments include things like defending one’s belief in discussion, revising it in the face of countervailing evidence, and allowing that belief to feature in an explanation of one’s behaviour in something like the following form: ‘I’m not voting for the Republicans because they are more likely to wage wars, I believe that wars are morally wrong, and I believe that one shouldn’t do things that further wrong actions’. And because beliefs have all these further commitments that you don’t want to take up if you merely *feel* a certain way about war, you will use an imperative to signal that you don’t want to take up all these commitments.

Error theorists should respond to this challenge by arguing that it is not a challenge. If imperatives express non-cognitive states and are therefore not truth-apt

---

<sup>488</sup> Joyce (2001: 176).

<sup>489</sup> Svoboda (2011: 43).

<sup>490</sup> Cf. Green (2009: §3) for an exception.



then they fall outside the scope of error theory and can be legitimately and felicitously uttered. And if imperative express beliefs with deontic, evaluative or thick moral content, then error theory kicks in. Overall, therefore, pervasiveness worries do not threaten error theory.

## 7.5 The Non-Arbitrariness Constraint and Metaphysical Possibility

I will now discuss:

### Coons' Argument

- P1 There is a metaphysically possible world that is non-morally identical to the actual world where error theory is false
- P2 If (P1) then there are moral facts in the actual world
- P3 There are moral facts in the actual world [From 1,2]
- P4 If (P3) then error theory is false
- C Therefore, error theory is false [From 3,4]<sup>491</sup>

I explain the objection and why it fails.

Start with P1. Since my error theory only argues that moral properties are metaphysically queer and not metaphysically impossible, there is a world like ours where error theory is false (i.e., where moral properties are instantiated). Coons argues that support for P1 comes from three common metaethical positions, accepted by error theorists like myself (and some other metaethicists as well):

- |               |   |
|---------------|---|
| First Claim:  | “no non-moral features of the actual world preclude the instantiation of moral properties” <sup>492</sup> |
| Second Claim: | moral claims are truth-apt  |
| Third Claim:  | moral claims are coherent (so we cannot determine on a priori grounds that they are false)                |

First Claim rules out the possibility that there is something about this world that guarantees that it can't instantiate moral properties even though moral properties are possible because they are instantiated in other possible worlds.<sup>493</sup> Second Claim rules out non-cognitivist interpretations of moral language which, at least without

---

<sup>491</sup> Coons (2011: 87-92). In fact Coons' argument is more complicated than this (it uses the wrongness of particular acts as well) but these complications don't affect the argument.

<sup>492</sup> Coons (2011: 87).

<sup>493</sup> Coons writes “This sort of skepticism is rare, but it is not outlandish. For example, imagine someone who argues that there are no moral requirements because no *actual* being has the cognitive capacities for agency, and, there are no moral requirements if no agent is bound by them” (2010: 87n10).

the quasi-realist agenda, could be pressed into the service of an argument for morality's impossibility.<sup>494</sup> Third Claim rules out the possibility of an error theory according to which moral properties are impossible because the concepts we use to pick them out make incompatible claims about their instantiation conditions.<sup>495</sup>

However, even if this is all true (hence, even if it is true that moral properties are instantiated in a world that is non-morally identical to ours), this doesn't yet show that moral properties are instantiated *in the actual world*. This is what P2 is meant to guarantee. Coons argues that the *non-arbitrariness constraint* on moral properties' instantiation conditions guarantees that moral properties would also be instantiated in the actual world once they are instantiated in a world that is non-morally identical to ours. According to the non-arbitrariness constraint, it is "impossible" that the instantiation of moral properties is "random and unprincipled", which is what it would be if "morality could require you to do one thing in a particular situation and something else in a qualitatively identical situation."<sup>496</sup> Coons also argues that a suitably strong or at least a global *supervenience relation* between moral and non-moral properties allows for the same inference.<sup>497</sup> In either case, the idea is that if moral properties are instantiated in a possible world non-morally indiscernible from ours, it then follows that, given the truth of the non-arbitrariness constraint or a suitable supervenience relation, moral properties are also instantiated in the actual world.

P4 is obviously true, and so is P3 as long as P1 and P2 are true. Moreover, the argument is valid. Therefore, Coons' Argument is a real challenge to my error theory, which was a claim about the actual world.

What to make of this argument? It is not clear that P1 is true. That is, it is not clear whether there is a possible world non-morally indiscernible from ours where error theory is false. If, as we are assuming, the actual world is inhospitable to moral properties because of its constellation of non-moral properties on which the moral properties supervene, then this fact precisely rules out the possibility that there is another possible world, non-morally identical to ours, where moral properties are

---

<sup>494</sup> See Brown (2013: 628); Wedgwood (2001: 3). I argue against the idea that emotivism entails morality's impossibility in Kalf (MS).

<sup>495</sup> For such error theories see Loeb (2008); Schiffer's (1990, 2003).

<sup>496</sup> Coons (2011: 91).

<sup>497</sup> Coons (2011: 90-1).

instantiated. I argued for the antecedent of this conditional in Chapter 6. As we also saw, Coons explicitly *assumes* the opposite with his assumption First Claim. But this is an extremely gratuitous assumption in this context.<sup>498</sup> Perhaps moral properties are possible, because instantiated, in other worlds *not* non-morally identical to ours. For perhaps in those worlds there are unicorns and irreducibly normative categorical moral reasons. But that doesn't show that moral properties are possible in other worlds that are non-morally *identical* to ours—they aren't, because in those worlds the same non-moral facts obtain, and these leave no room for moral properties, as I argued in Chapter 6. Coons' Argument fails.<sup>499</sup>

## 7.6 Conclusion and Preview

In this chapter I have argued that the remaining objections to error theory that can be found in the literature are not convincing. Given my earlier argumentation in Chapters 1-6, we should therefore conclude that error theory is a plausible metaethical theory—indeed, that it is the most plausible metaethical theory within a cognitivist framework. The next chapter formulates and defends revolutionary cognitivism as a response to error theory.

---

<sup>498</sup> Also see Brown (2013).

<sup>499</sup> This also means that my error theory is *sufficiently stable*. We can't seem to change something about the fact that our world is one in which metaphysical naturalism holds. So there doesn't seem to be anything we can do to 'introduce' morality in the actual world. Had morality depended not on facts that can't be reconciled with metaphysical naturalism but on facts that *can* be reconciled with metaphysical naturalism and that are just contingently not instantiated (e.g., identity of morally relevant desires across all agents—Burgess 1998) then we would have been able to introduce morality into the actual world. All that is needed for that is for us to adopt the same morally relevant desires. But, again, for my error theory this isn't possible.

## Chapter 8

# Revolutionary Cognitivism

### 8.1 Introduction

Thus far I have argued that error theory, or at least my formulation of it, merits more credit than it is often given. Indeed, my arguments suggest that we should be error theorists rather than success theorists if we are metaethical cognitivists. This raises the following question: what (prudentially) should we do with our error-ridden moral discourse? I argue that adopting a hitherto largely underexplored version of revolutionary cognitivism is the best answer to that question.<sup>500</sup>

In the next section I explain what revolutionary cognitivism is. I discuss its psychological, semantic and epistemological commitments. I also explain its benefits (§8.2). After this I discuss problems for revolutionary cognitivism and suggest solutions to these problems (§8.3). The combined force of §8.2-3 is that revolutionary cognitivism is plausible. I then compare revolutionary cognitivism with other responses to error theory, arguing that revolutionary cognitivism enjoys important benefits over each of them in pair-wise comparisons (§8.4). The upshot of §8.2-4 is that we should become revolutionary cognitivists after error theory. A conclusion summarizes my main findings, both of this chapter and that of the thesis as a whole (§8.5).

### 8.2 Psychology, Semantics, Epistemology, Benefits

*Psychology.* Revolutionary cognitivists believe that after error theory we should stop believing moral propositions and should start believing *schmoral* propositions, where schmoral propositions are propositions like:

---

<sup>500</sup> ‘Hitherto largely unexplored’ because although revolutionary cognitivism is, as I argued in Chapter 1, *Mackie’s* response to the truth of error theory, his response to error theory has been largely misunderstood. Moreover, in more recent times revolutionary cognitivism is not even mentioned as an option for error theorists in for example Miller (2013: 115); Joyce (2001, 2005); Suikkanen (2013). Two exceptions are Schiffer (1990, 2003) and Lillehammer (1999).

**Schwrong**      Stealing is schmorally wrong

The fact that revolutionary cognitivists continue to talk about beliefs sets them apart from certain types of fictionalist views that say that we should alter our mental attitudes towards moral propositions. Recall the attitude-fictionalist from §3.5. Schematically:

Attitude of make-believe to:	Moral proposition (E.g., ‘stealing is morally wrong’)
------------------------------	---

Other fictionalist views propose to keep our attitude towards moral propositions as it is now (i.e., one of believe) and suggest that we should alter the content of moral propositions by prefixing them with a fictionalist operator.<sup>501</sup> These are content-fictionalist views. Schematically:

Attitude of believe to:	Fictionalist moral proposition (E.g., ‘in the moral fiction, stealing is wrong’)
-------------------------	--

Revolutionary cognitivism is similar to content-fictionalism in that it keeps intact the attitude of believe we currently bear toward moral propositions and only alters the content of these moral propositions. But according to the revolutionary cognitivist, we shouldn’t prefix moral propositions with ‘in the fiction’-operators. Instead, we should change the content of what we say by starting to talk about *schmoral* rather than moral wrongs:

Attitude of believe to:	Schmoral proposition (E.g., ‘stealing is schmorally wrong’)
-------------------------	---

*Semantics.* What does it mean for stealing to be *schmorally* wrong? It means that there are considerations that speak against stealing that fall short of being *moral* considerations. Recall that a consideration counts as a *moral* consideration if and only if it involves *categorical normativity in the rationality-involving sense*. If one utters a moral judgment like Wrong then one’s utterance is false if there isn’t any categorical normativity in the rational requirement-involving sense not to perform the action that one said was wrong. Contrastingly, if one utters a *schmoral* judgment like Schwrong then one’s utterance is false if there wasn’t any *hypothetical* normativity in the rationality-involving sense not to perform the action that one says

---

<sup>501</sup> Oddie and Demetriou (2010: 200).

is Schwrong. Hypothetical normativity in the rationality-involving sense is normativity that governs combinations of agents' mental states and is grounded in agents' desires. In Chapter 6 I argued that there is no categorical normativity in the rationality-involving sense that is recognisably moral. But that leaves it open that there is hypothetical normativity in the rationality-involving sense that is recognisably moral. Call considerations that have hypothetical normativity in the rationality-involving sense and that are recognisably moral *schmoral* rather than moral considerations. Revolutionary cognitivism proposes that we start uttering schmoral propositions that describe hypothetical normativity in the rationality-involving sense.

What is hypothetical schmoral normativity? Consider *non-substantive, procedural* rational requirements such as the Instrumental Principle (IP). IP can be used to argue that agents *who already desire not to steal* should not steal *this particular bike*. Here the reasoning is that (i) being in the mental state of knowing that stealing this bike would not further one's more general, underived desire not to steal and that (ii) being in that general mental state of not desiring to steal generates, under normative pressure of IP, (iii) the further desire not to steal this particular bike. But again, although we do have normativity in the rationality-involving sense with recognizable moral content ('do not steal this bike'), we don't have *moral* normativity. For the normativity we have is *desire-dependent or hypothetical*—drop the general desire not to steal and the normative pressure you were under not to steal *this bike* evaporates. Of course IP itself is categorically normative in the rationality-involving sense. But IP has no content—IP is a purely procedural rational requirement. Only in conjunction with (i) and (ii) does it generate normative pressure for the agent to form the instrumental desire not to steal the contextually salient bike. It is for this reason—i.e., that we can only get rationally authoritative normative pressure on agents to adopt desires with recognizable moral content if we combine a procedural rational requirement with a non-instrumental, recognisably moral *and yet rationally optional* desire that we can drop whenever we want to escape the rationally authoritative normative pressure—that I call the normativity that does exist and that is recognisably moral 'hypothetical' normativity. As explained in Chapter 5, the folk would not mistake their categorical moral normativity with this kind of hypothetical normativity. But since we are talking

about *schmorality*, this is fine—and, indeed, the best we can have given the sceptical arguments in Chapter 6.

A final remark about the semantics of revolutionary cognitivism is this. What the revolutionary cognitivist proposes is a *homophonic translation* for moral terms in moral judgments to get schmoral terms that feature in schmoral judgments:

what our utterances of [schmoral] sentences ... reflect are our own subjective values. So ... [we might] carry on as before [with moral discourse] but without imagining our evaluations to be [categorically normative in the rationality-involving sense]. If we would then be uttering sentences which are capable of being true, what would *make* them so ... is fidelity to our subjective values<sup>502</sup>

The translation is homophonic because although the content of what is being said is altered, what the resulting utterance sounds like is not different from the original utterance. The folk, argues the revolutionary cognitivist, can continue to *say* things like ‘stealing is wrong’ as long as what they mean (I’m talking about semantic rather than speaker meaning here) by that is ‘there is *hypothetical* normativity in the rationality-involving sense that governs my desires-forming processes’. Metaethicists discussing the pros and cons of revolutionary cognitivism are probably better advised to use the nomenclature of ‘morality’ and ‘schmorality’, however, in order to avoid confusion.

*Epistemology.* How do we know what is schmorally right and wrong? We can know this by looking at our own subjective values. If I have an original desire to steal and no desire not to steal then, given the Instrumental Principle and as explained above, I should steal. Contrastingly, if I have a desire not to steal and no desire to steal then, again given the Instrumental Principle and as explained, I should not to steal.

---

<sup>502</sup> Burgess (1998: 545; emphasis in original). I borrow the term homophonic translation from Burgess. Also see Lillehammer (2000: 174) who suggests that error theorists could say that “normative reasons *should* be construed as response dependent regardless of the conceptual commitments embodied in common sense ethical discourse.” Similarly Balaguer explains that after error theory we “might decide that we would be better off if we altered our moral practices, i.e., if we started using our moral terms slightly differently, so that they expressed slightly different concepts, concepts that were ... more natural or coherent, or some such thing” (2011: 374).

A major question is whether, and if so how, our desires should be corrected in light of full or ideal information or rationality constraints.<sup>503</sup> My revolutionary cognitivism holds that some sort of correction of desires vis-à-vis other desires is mandatory. If I have the fundamental desire not to steal but also have the more specific desire to steal *this bike*, then one of these desires has to go. On the safe assumption that agents will more often relinquish their particular desires rather than their general desires, we can assume that in this situation the agent will relinquish her desire to steal the contextually salient bike. But since the Instrumental Principle, which is supplying the normativity in this case is content-neutral or a *procedural* rather than a *substantive* rational requirement, it would be equally fine for the agent to drop her fundamental desire for not stealing and to keep the desire to steal the bike (and then probably also to adopt the new fundamental desire *to steal*). So according to revolutionary cognitivism correction of desires vis-à-vis other desires is mandatory, but correction of desires *per se* is out. For that requires substantive rational requirements with recognisable moral content, and these don't exist. Agents are not doing anything wrong when they have the fundamental, non-derivative desire to steal. They are certainly not doing anything *morally* wrong.<sup>504</sup>

*Benefits.* Why schmoralize? The basic idea is that there is a subset of the roles that morality currently plays that could be played by something *like* morality even if it became widely accepted that morality is flawed.<sup>505</sup> There are two aspects of moralizing that schmoralizing can mimic and that constitute schmorality's benefits. One aspect of morality that schmorality can mimic is morality's role in *intra*-personal practical deliberation. The second aspect of morality that schmorality can mimic is morality's role in *inter*-personal moral communication. Let's start with intra-personal benefits.

---

<sup>503</sup> The issue is a very thorny one as well (Enoch 2005).

<sup>504</sup> It would indeed be implausible if we were "[trimming moral] obligation to the size of individual motives" (Frankena 1958: 80; also see Brandt 1979) but we aren't doing that. In the wake of error theory we are proposing that we are trimming *schmoral* obligations to the size of individual motives.

<sup>505</sup> West (2010: 196); Lillehammer (2004: 100); Parfit (2012: Vol. I, p273).



*Intra-Personal Benefits.* We moralize, and might have to continue to schmoralize, about things we deeply care about.<sup>506</sup> Moralizing was, and schmoralizing will be, about torturing, killing, stealing, cheating, lying, giving to famine relief, helping someone at great costs to oneself, scoffing at the vulnerable, abortion, euthanasia, starting wars, sending more soldiers to the front, ending wars, giving kidneys to family members, *et cetera, et cetera*. The revolutionary cognitivist argues that there is normative pressure not to engage in some of these activities and to engage in some of the others. True, the normative force that it countenances stems from procedural rational requirements like the Instrumental Principle combined with our contingent desires rather than from an independently existing and irreducibly normative moral reality or substantive rational requirements that supply the required normative force. But that difference is often unimportant. It is a safe conjecture that many healthy, responsible adults desire not to steal and not to kill, and do desire to give to famine relief.<sup>507</sup> One benefit of revolutionary cognitivism is that with it we can still make sense of the thought that we have to do these things. After all, we desire to give to famine relief and believe that we can do this by donating money to Oxfam. Then, by the Instrumental Principle, we are under genuine normative pressure to form a specific desire to donate to Oxfam. Indeed, not only can we, with schmorality, make sense of the thought that people with the right desires should give to famine relief, we can also make sense of the thought that such adults are making a rational *mistake* in their practical deliberation if they don't settle on giving to Oxfam under these circumstances. For it is a rational mistake if one doesn't desire to give to Oxfam and yet desires to give to famine relief and believes that giving to Oxfam is a good enough (or the best, or the most economic) way of giving to famine relief.

So the first benefit of schmoralizing is this

**First Intra-Personal Benefit of Schmoralizing**

Schmoral considerations constrain agents' practical deliberations almost as much as moral considerations do, thereby allowing us to hold on to the ideas that we should not desire to steal bikes (under certain circumstances, including the obtaining of the

---

<sup>506</sup> I write 'might have to continue to schmoralize' because at this point in the chapter I cannot yet rule out other options such as fictionalism and preservatism. I argue that revolutionary cognitivism outcompetes these and other theories in §8.4.

<sup>507</sup> This even holds for nihilists as their "past beliefs may have continuing effects on what they care about and do" (Parfit 2012 Vol2: 462). Also see Burgess (1998).

relevant desires) and that we are making a rational mistake when we do steal (again under certain circumstances, including the obtaining of the relevant desires).<sup>508</sup>

‘Almost just as much’ because what schmorality can’t make sense of is the thought that *all* adults have to abide by particular rules and obligations with a particular content, such as the rule against stealing, whatever they desire. But that is precisely what necessitated the error theory—there are no rational requirements governing the formation of underived desires with recognizable moral content. Other than this there are a lot of other things that the moralist can say and that the revolutionary realist can say just as well; including, as indicated, that *most* agents are *most of the time* making rational mistakes if they don’t give to famine relief (if the right background desires and beliefs are in place). ‘Most agents’, because there are going to be some agents who simply desire to steal and that’s that. And ‘most of the time’ because even though for many agents it is true that they ought to drop their instrumental desire to steal a particular bike when they don’t want to drop their fundamental desire not to steal, it is perfectly legitimate for those agents to decide to drop their fundamental desire not to steal instead. But these exceptions notwithstanding, again, for the majority of cases revolutionary cognitivism gives the same verdict as ordinary morality—viz., that agents ought to refrain from stealing and that they therefore ought to drop their motivationally efficacious desire to steal.

Does schmoralizing have to be as superficial as I have presented it so far? That is, can schmoralizing only consist in figuring out something as straightforward as whether an act of giving to Oxfam strokes with one’s beliefs and desire to give to famine relief?

The answer is ‘no’, and this is nicely brought out by Lillehammer. He starts by saying much what I say about how to respond to error theory:

the falsity of [moral] judgments does not entail the falsity of judgments which assert the existence of instrumental relationships between means and ends. Nor does the falsity of [moral] judgments entail the falsity of judgments which say whether or not different practical options are consistent with each other, given certain constraints or given certain circumstances<sup>509</sup>

He adds to this picture that:

---

<sup>508</sup> Lillehammer (2004: 108); also see Railton (1986: 169); Hume (1751: Sect. IX, Pt. II).

<sup>509</sup> Lillehammer (1999: 211); also see Timmons (1999: 82).

There is an indefinite number of truths about how ends can be promoted and related to each other, necessarily or contingently. The [error theorist] can invoke these facts to explain why informative, insightful, deep and useful ethical thinking can be undertaken even if there are no [moral] normative reasons. If the ends involved are of the requisite kind and the problem situation sufficiently complex, the thinking required to make sense of the situation can easily amount to highly systematic and illuminating moral theorizing<sup>510</sup>

One example of such ‘systematic and illuminating moral theorizing’ concerns the question of the improvement of the human condition. How is that to be attained? That is, how do factors such as world-politics, poverty, the power of multi-nationals and the current economic crisis affect this issue—and what overarching schmoral principle does it make sense to desire given one’s current set of both non-instrumental and instrumental desires and beliefs? Schmoralizing is not always as simple and straightforward as my examples involving IP have so far suggested. Moreover,

[schmoral] theorizing can also provide interesting impossibility and possibility proofs based on either the logical consistency or the practical compatibility of different conceptions of the good or the right. For example, one might try to prove that a system of human rights can be implemented consistently with a utilitarian moral theory ... All the [error theorist] denies is that [a certain] approach [to the question of our ultimate ends] is intrinsically rationally privileged<sup>511</sup>

Again, schmoralizing will not, and cannot, be identical to moralizing. But it can be something that mimics moralizing to a fairly large degree and can it can be a very interesting and difficult activity all the same. It also follows from this that revolutionary cognitivism can make sense of moral disagreement, as there can be deep and interesting disagreement about what the (correct) means-end relations are as they pertain to a particular moral problem and its solution.

But why not think *directly* in terms of desire-satisfaction rather than in terms of what is ‘schmorally’ right or wrong? This is where the second, intra-personal benefit of revolutionary cognitivism becomes relevant. It is generally true that the more cognitive steps are required in thinking about something, or in making a decision, the less likely it is that the agent reaches a particular conclusion or acts on a particular intention.<sup>512</sup> The more cognitive steps one needs to make the more likely it is that where one wants to end up—in our case, the situation in which one has a

---

<sup>510</sup> Lillehammer (1999: 211).

<sup>511</sup> Lillehammer (1999: 213).

<sup>512</sup> Tresan (2010: 230); cf. Mackie (1977: 156); Joyce (2001: 184, 212).

psychology that maximally adheres to the rational requirements—isn't where one will in fact end up. Once the agent knows that stealing is inconsistent with her underived desires and instrumental beliefs and so long as there aren't any signs that anything has changed with her attitudes towards stealing, then the following holds. Thinking directly in terms of the schmoral wrongness of stealing allows the agent to skip some cognitive steps and to form a natural bulwark against stealing, adopting the desire for which allows the agent to attain a psychology that is fully in line with the rational requirements—or at least more in line with the rational requirements than a psychology in which she (wrongly) instrumentally desires to steal the bike. So the second benefit of schmoralizing is this:

**Second Intra-Personal Benefit of Schmoralizing**

Agents using schmoral considerations to govern their attitude-formation processes can better, because more reliably, approach a psychology that is maximally in line with the existing rational requirements

To sum up the discussion so far, revolutionary cognitivism undertakes various psychological, semantic and epistemological commitments. These commitments come with benefits. Two of these are intra-personal benefits; benefits that concerns the agent insofar as she herself uses schmoral considerations in her practical deliberations. One such benefit is that with schmoral wrongs and obligations we can continue to make sense of the thought that we *have to* muster desires and intentions for performing or not performing morally relevant actions such as killing and stealing and indeed that we can make sense of the thought that we are making *a rational mistake* if we are not mustering those desires. The other benefit is that thinking in terms of schmoral wrongness rather than directly in terms of what is consistent with one's other mental attitudes and the (procedural) rational requirements probably gives the agent a better chance of getting a psychology that maximally approaches the ideal of one that adheres to all the rational requirements. Let me now consider the inter-personal benefits of schmoralizing.

*Inter-Personal Benefits.* There are also inter-personal benefits associated with revolutionary cognitivism. Two of these mimic the two intra-personal benefits just discussed. I discuss them first and then turn to a third interpersonal benefit.

It is an *intra*-personal benefit of schmoralizing that the agent can make sense of her own obligation to refrain from stealing and indeed that she can make sense of her making a rational mistake when she does steal (so long as she has some relevant desires grounding the truth of these claims). Similarly, it is an *inter*-personal benefit

of schmoralizing that we can make sense of *other people* having to refrain from stealing and we can also make sense of them making a rational mistake when *they* steal (as long as they have some relevant desires grounding the truth of these claims). Let me elaborate on both conjuncts of the inter-personal claim.

First, we can make good sense of other people having to refrain from stealing if we accept revolutionary cognitivism because according to that theory there are rational requirements governing individuals' combinations of mental states such that when agents desire not to steal and know that taking away this bike amounts to stealing, they have to muster a desire for not-stealing that bike. On the assumption that agents should not drop true beliefs and on the further assumption that they will be reluctant to drop their original, non-derivative desire for not stealing, the attitude that is likely to be dropped and should be dropped if the assumption is true will be the one for stealing the bike.

Second, if agents don't make any of these changes then they are making a rational mistake, and revolutionary cognitivism therefore allows us to direct legitimate *critique* of their behaviour at other people from the point of practical rationality. The critique is 'legitimate' because it reflects the fact that these agents are ignoring normative force in the rationality-involving sense to perform certain actions.<sup>513</sup> Of course, the kind of critique can be more interesting and difficult to put one's finger on than this (recall my discussion of Lillehammer's end-relative moral theorizing above), but these are the essentials of the first inter-personal benefit of schmoralizing. So we get:

**First Inter-Personal Benefit of Schmoralizing**

Schmoral considerations constrain agents' practical deliberations almost as much as moral considerations do, thereby allowing us to hold on to the ideas that *other agents* should not desire to steal bikes (under certain circumstances, including the obtaining of the relevant desires) and are rightly rationally criticisable when they do steal (again under certain circumstances, including the obtaining of the relevant desires)

---

<sup>513</sup> Whether agents are culpable for not knowing that this force exists in addition to being culpable for knowing that the force exists and yet don't act on it is not immediately relevant. After all, with the latter kind of culpability we can already provide a vindicatory explanation of our intuitions that other agents ought not to steal—at least when assuming that these agents are normal, healthy human beings and that they are also unlikely to relinquish their non-instrumental, underived desire for not-stealing.

A second inter-personal benefit of schmoralizing mimics the other intra-personal benefit of schmoralizing. The second benefit of schmoralizing on the *intra*-personal level was that by taking certain considerations as schmoral considerations and simply acting on them (rather than going through the underlying calculations of what accords with one's fundamental desires in combination with the rational requirements, *et cetera*) will make it easier for the agent to attain the ideal of a psychology that is maximally adherent to the rational requirements. A similar benefit with schmoralizing materializes on the inter-personal level. The thought is that in order to get what you and I both want (this time a world that contains a little suffering as possible, on the assumption that that is what we both fundamentally desire) it will be useful to be able to skip some cognitive steps. We will have more success attaining the world that we both want when we can say to each other that stealing is schmorally wrong compared to when we communicate our ideas in the following, highly convoluted way: 'stealing is inconsistent with our shared non-derivative desire of living in a world in which there is as little human suffering as possible because and ... etc., etc.'<sup>514</sup> So we get:

#### **Second Inter-Personal Benefit of Schmoralizing**

Schmoral considerations allow agents to better communicate and thereby better attain the world that is most congruent with their shared desires compared to a situation in which those agents cannot communicate such considerations

In addition to these two inter-personal benefits there is one further such benefit.<sup>515</sup> Distinguish between the locutionary and perlocutionary act of schmoral

---

<sup>514</sup> Similarly Schiffer remarks that schmoralizing allows for "quick and easy" coordination of plans among "certain recognizable others about the kind of world we wished to inhabit" (2003: 261). Schiffer also writes that perhaps "we would introduce an indexical word 'shmwrong' such that when A said to B that it would be shmwrong for so-and-so to do such-and-such she was expressing a belief whose case-value was that A and B would want so-and-so not to do such-and-such if they were agreed about all relevant facts ... My use [of moral language] is instrumental, but it gets across what I need to get across, and which I couldn't begin to get across if I had to express the truth in what I was saying without use of those terms" (2003: 261). I actually don't think that it is true that we 'couldn't begin to get across' these behaviour-cooperation plans without a moral or schmoral language, but other than that I am in broad agreement with Schiffer's remarks.

<sup>515</sup> It could be thought that there is a fourth benefit—viz., that with schmoral discourse we can make sure that people continue to behave in ways that we used to think were morally obligatory or morally laudable. This was Mackie's proposal—we should schmoralize *in order to* ensure that we can reap the benefits of social

utterances, which are both speech acts. The locutionary act one performs in uttering a (schmoral) judgment is the act of ““saying something” in the “full normal sense.””<sup>516</sup> So when I say ‘stealing is schmorally wrong’ the locutionary act I perform is that of saying that stealing is schmorally wrong. But in uttering the same sentence I can also perform the perlocutionary act of *scaring* someone or *bullying someone into believing that stealing is wrong*. What perlocutionary act I perform in uttering ‘stealing is wrong’ is—and this contrasts with locutionary acts—at least in part dependent on the hearer’s reaction. Whether I scare John by uttering ‘stealing is schmorally wrong’ depends in part on how he interprets my utterance (but probably also at least in part on how I utter that judgement). Likewise, whether I can bully John into believing that he shouldn’t steal depends on our past relation or his perception of my power, or something relevantly similar.<sup>517</sup>

I use the word ‘bullying’ in reference to Williams’ thoughts on external reason claims.<sup>518</sup> These are claims about what agents have reason to do for which it is not part of their truth-conditions that the addressee has a desire (or can reach that desire through certain procedurally rational modifications of her existing set of desires ‘S’). It was precisely Williams’ view that external reason claims aren’t anything other than attempts to *bully* agents into doing something. After all, there is nothing about the addressee that resonates with an external reason claim. The addresser simply tries to get the addressee to do something for which, according to Williams anyway, there is no justification. We have ‘bullying’, not ‘reasoning’ or something akin to reasoning.

Why would it be a benefit of revolutionary cognitivism that agents can be ‘bullied’ in this sense? The benefit is that schmoral talk enables agents to try to influence the behaviour of others with whom they have a fundamental moral disagreement (i.e., a disagreement about underived, basic desires such as desires for the world to contain or not to contain as much suffering as possible). Given that the stakes are high—the realization of one’s ideal world in which (it seems safe to hope) stealing is wrong—it is a good thing that we can use schmoral language for the

---

cooperation. But we also saw that this proposal won’t work (§2.2.4). Better than to drop the idea that schmorality is useful for the attaining of some social end or goal, and to rely on the benefits that it can generate.

<sup>516</sup> Austin (1975: 94).

<sup>517</sup> For similar ideas see Lillehammer (1999: 205).

<sup>518</sup> Williams (1979); also see his (1995a, 1995b, 2001).

purpose of bullying other agents into believing what we believe. Agents who believe that the world should contain as little suffering as possible perhaps won't want to brutally force other agents to change their ways. But they might feel comfortable enough to 'bully' those agents to adopt their desires. Furthermore, bullying is not *morally* wrong because nothing is morally wrong. So the worry that this aspect of revolutionary cognitivism is *morally* wrong, I remind the reader, is a non-problem. Thus we also have:

**Third Inter-Personal Benefit of Schmoralizing**

Schmoral considerations can be used to get other agents to share one's fundamental moral point of view

Such are the benefits of revolutionary cognitivism. I will now turn to its (potential) drawbacks and my arguments that in fact these are not drawbacks.

### **8.3 Objections and Replies**

My discussion of revolutionary cognitivism raises various questions:

- (1) Does revolutionary cognitivism collapse into moral realism?
- (2) Is implementing revolutionary cognitivism practically feasible?
- (3) Is revolutionary cognitivism consistent with the most plausible account of the nature of belief?

In what follows I aim to answer these questions. If I'm successful in this endeavour and combining this with the result of the previous section—viz., that revolutionary cognitivism has five benefits—we'll have a strong, but defeasible case for revolutionary realism as our answer to the question 'what (prudentially) should we do with moral discourse after error theory?' That this case isn't in fact defeated by the benefits (if indeed there really are any) of the other proposals, including fictionalism, will be argued in §8.4.

#### **8.3.1 A Collapse Into Realism?**

Jon Tresan asks, once we have made the change from uttering moral propositions to schmoral propositions, why



we haven't thereby created new moral terms. After all, the substitutions [i.e., the schmoral predicates] would pick out the very properties naturalists say are the moral properties<sup>519</sup>

Tresan's problem is that error theory *coupled with a revolutionary cognitivist procedure for assigning moral terms new referents*—viz., action tokens and types that there is hypothetical normativity in the rationality-involving sense to perform or to refrain from performing—seems to collapse into a form of moral *realism* or a *success theory*. After all, the quote suggests, the properties we pick out with schmoral and moral terms—actions of stealing and the like—are the same and they are both normative in the rationality-involving sense. Their only difference is that moral normativity is categorical and that schmoral normativity is hypothetical. Tresan claims that this difference is enough to justify the claim that schmorality collapses into morality.

Error theorists supporting revolutionary cognitivism can respond to this objection in two ways. First, they can insist that this difference between morality and schmorality *is* enough to keep the two apart. Indeed, that was part of the message of Chapter 5. The folk simply wouldn't mistake their categorical moral normativity for hypothetical schmoral normativity. Second, error theorists can argue that in fact the revolutionary cognitivist is not committed to the view that the 'substitution schmoral predicates would pick out the very properties naturalists say are the moral properties.' After all, if I am under hypothetical schmoral normativity either to drop my instrumental desire not to steal this bike or my non-instrumental desire to steal then according to my version of revolutionary cognitivism, agents are acting fully rationally when they drop the instrumental desire not to steal and replace it with a different instrumental desire to steal the bike. But ordinary morality wouldn't allow this. So that's another, important difference between moralizing and schmoralizing. This objection doesn't work.

### 8.3.2 Feasibility

For a proposal like revolutionary cognitivism it is important that it is practically feasible. We can accept as a general methodological maxim in practical philosophy that the unfeasibility of a thesis diminishes the credence we should have in it. For an example from normative moral philosophy, take a theory that imposes a duty to

---

<sup>519</sup> Tresan (2010: 236); also see Lillehammer (1999: 215); Joyce (2007b: 69); Lewis (1989: 93).

practice impartial benevolence among all people such that we can show no greater concern for the welfare of ourselves and loved ones than we do towards billions of others. That theory is probably unfeasible.<sup>520</sup> The evolutionary history of human beings suggests that they won't be able to refrain from favouring kin over non-kin in certain circumstances.<sup>521</sup> Therefore, all else being equal, this theory is less likely to be a true theory in *practical* philosophy than a similar theory that imposes a more feasible duty. True, if a theory is difficult to implement than that need not imply that the theory is false. But if a theory is (virtually) impossible to implement, perhaps because it goes against the essentials of human nature, then that does seem to imply that the theory is false. After all, the whole point of doing *practical* philosophy is that one's proposals are, precisely, practically relevant rather than interesting but unfeasible.

The worry for my view is that something similar might be going on with revolutionary cognitivism. If revolutionary cognitivism is unfeasible then, all else being equal, we should diminish our credence in it compared to rival theories that are feasible. And revolutionary cognitivism might be thought to be unfeasible for two reasons. First, it seems difficult to convince the folk of error theory and revolutionary cognitivism. Understanding error theory requires a high level of conceptual sophistication that the folk cannot be assumed to have, argue my opponents.<sup>522</sup> Second, even if we can teach the folk enough metaethics to understand error theory we still face the worry that we might not have the means to do this. As Cuneo and Christy put it:

how would one go about communicating [the error-theoretic] message to the world (late night commercials?)<sup>523</sup>

Both problems can be solved in one go. The solution is to allow the vast majority of people who, at least by hypothesis, cannot understand the error theory *to remain realists*. This solves the first problem because if we can allow a lot of people to remain realists then the difficulty of how we are going to *convince* those people of the truth of error theory (and revolutionary cognitivism) disappears. And this solves

---

<sup>520</sup> Freeman (2012).

<sup>521</sup> Joyce (2006, 2001: Chapter 6).

<sup>522</sup> Cuneo and Christy (2011: 94). Also see Oddie and Demetriou (2010).

<sup>523</sup> Cuneo and Christy (2011: 95).

the second problem because if we can allow a lot of people to remain realists then the difficulty of how we are going to *reach* everyone to preach the error-theoretic message will then also disappear.

So we get a *mixed-world* of realists and revolutionary cognitivist. Why is a *mixed world* with moralizers and schmoralizers good enough—why shouldn't we strive for a *pure world* that only contains schmoralizers?

Part of the answer to that question is that there is at least nothing that speaks *against* having a mixed world. First, there is nothing *morally* wrong with a mixed world. We don't have a *moral* obligation to tell people the truth because we don't have any moral obligations at all. But Richard Joyce—who himself prefers an attitude-fictionalist treatment of our error-riddled moral discourse—argues that keeping the truth away from the folk results in an unstable society that it would be a *prudentially* bad society to live in.<sup>524</sup> If the truth comes out then chaos ensues because people will not know what to do or what to believe. They will also feel that they've been misled and deceived, with all the (prudentially) bad consequences that come from the uproar that might ensue in response to these feelings. Apart from the fact that this last point is an unwarranted empirical speculation, my position isn't subject to Joyce's objection because I don't argue that we should keep the truth away from the folk. We should be open and transparent about error theory and allow to folk to believe it if they can. We should not hide the truth from them. And if we don't hide anything from the folk then the kind of chaos that Joyce predicts does not seem to be forthcoming. Moreover, in a mixed-world we can have all the advantages of revolutionary cognitivism as discussed in this chapter, albeit on a smaller scale. So neither version of the unfeasibility objection (we can't convince the folk of error theory, and we can't reach them to try to convince them anyway) is viable.

Before closing however I want to point out that the claim that the folk won't be able to understand error theory is probably a false empirical speculation. Cuneo and Christy surmise that the folk won't be able to understand error theory due to the high level of conceptual sophistication that is required to understand such arguments as Blackburn's objection from supervenience against moral realism and Horgan and

---

<sup>524</sup> Joyce (2001: 214).

Timmons' moral twin earth argument.<sup>525</sup> But the error theory has a much simpler two-pronged form: here is a commitment of moral discourse that makes it what it is, and here is why it is false. So *if* we are speculating, it seems to me that the folk would be able to understand error theory.

### 8.3.3 Doxastic Involuntarism

The third and final objection to revolutionary cognitivism that I can think of is what I call the objection from Doxastic Involuntarism.<sup>526</sup> Error theory is committed to:

<b>Evidence</b>	If error theory is true then there is sufficient evidence for the falsehood of our first-order atomic moral beliefs
-----------------	---

In fact, error theory is normally taken to be committed to something stronger than Evidence, viz.:

<b>Falsity</b>	If error theory is true then all first-order positive atomic moral beliefs are false
----------------	--

I will get back to the relation between Evidence and Falsity shortly. First I remark that error theorists are also committed to:

<b>Continue</b>	It is possible to continue to have moral beliefs whilst believing error theory
-----------------	--

Most error theorists are committed to Continue because they argue that there are *prudential* normative reasons to abolish or make-believe (or what have you) moral thought. But it only makes sense to say that we should abolish or make-believe (or what have you) moral beliefs if it is possible to continue to hold on to them. Had it been impossible to continue to hold on to our error-riddled moral beliefs than arguing that we shouldn't is futile. As revolutionary cognitivism gives prudential reasons for engaging in schmoral discourse it must commit itself to the possibility of continuing to have moral beliefs in the wake of error theory.

The fact that error theorists both accept Falsity and Continue is problematic.<sup>527</sup> Falsity and Continue are logically inconsistent. You can't both accept Continue by believing a moral proposition like Wrong and accept Falsity at

---

<sup>525</sup> Cuneo and Christy (2011: 94). See Blackburn (1971); Horgan and Timmons (1992a, 1992b).

<sup>526</sup> Suikkanen (2013).

<sup>527</sup> Also see Streumer (forthcoming: §3).

the same time. For if you do that then you are both believing that stealing is wrong is true and that stealing is wrong is false.

To avoid this predicament Suikkanen allows error theorists to run their theory with the weaker thesis Evidence.<sup>528</sup> It is not logically inconsistent to believe Evidence and Continue at the same time. For it is not logically inconsistent to believe that stealing is wrong is true and that there is sufficient evidence that stealing is wrong is false at the same time. It may be epistemically ill-advisable to believe this, but it is not flatly logically inconsistent.

However, now consider:

**Doxastic Involuntarism**

It is constitutive of belief that it is impossible to choose to continue to believe what one believes one has sufficient evidence to believe to be false

According to Doxastic Involuntarism, if you have what you take to be sufficient evidence that grass is green, you can't continue to believe that grass is red.

The problem for any view about how to respond to error theory is that Doxastic Involuntarism renders the combination of Evidence and Continue worse than it already was. For Evidence and Continue together entail that moral beliefs are *systematically truth-insensitive attitudes*: attitudes that are not to be given up when one believes that there is sufficient evidence against them. That, after all, is what made it possible to accept both Evidence and Continue. But this is inconsistent with Doxastic Involuntarism, according to which beliefs are *systematically truth-sensitive attitudes*. Something has to go.

Indeed, the inconsistent triad of Evidence, Continue and Doxastic Involuntarism generates a trilemma. Horn one: Either drop Evidence, but that is to drop (belief in) error theory. Horn two: drop Continue, but that necessitates dropping the quest for a normative theory for moral discourse after error theory and to engage in empirical speculations about *what will in fact happen* with moral discourse if we were to become error theorists. Examples of such descriptive theories are descriptive abolitionism and descriptive preservatism, according to which we will as a matter of fact abolish or preserve moral thought and talk in the wake of error theory. The problem with this is that such empirical speculation is very uninteresting. Horn three: drop Doxastic Involuntarism, but that incurs the obligation to adopt and

---

<sup>528</sup> Suikkanen (2013).

defend a non-standard account of the nature of belief. The objection to Revolutionary Cognitivism, since it offers prudential reasons for schmoralizing rather than continuing to moralize, is that it is committed to this inconsistent triad of claims.

In response, error theorists should accept the second horn of the trilemma. For it is not true that denying Continue obliges one to recognize that our only options are to wait and see which of the descriptive theories turns out to be correct. Even if we can't hold on to our old moral beliefs and as a matter of fact automatically abolish all moral thought and talk after error theory, it is still an interesting question whether we *should* do something else—and my own answer is that we should start to schmoralize.

My solution to Suikkanen problem can be applied, *mutatis mutandis*, to the other theories about what to do after error theory. Revolutionary cognitivism enjoys no special advantage over any of them insofar as the Problem of Doxastic Involuntarism is concerned. However I will now argue that revolutionary cognitivism is to be preferred over each and all of these.

## **8.4 Comparing Revolutionary Cognitivism with Other Views**

In this section I compare revolutionary cognitivism with other views about what is the (prudentially) best response to the truth of the kind of error theory that I have been defending in this thesis.

### **8.4.1 Abolitionism**

Abolitionists argue that we should abolish moral thought and talk, usually because this is prudentially better for our overall well-being.<sup>529</sup> The best example I know to support this claim comes from the euthanasia and abortion debates. The abolitionist claims that in the absence of moral thought and talk we can focus in the relevant debates on the well-being of all parties involved. This is then argued to have hugely beneficial effects: by focussing directly on human well-being and by deliberately ignoring complicating and constraining moral factors we should be able to become much better at improving the human condition.

One problem for the abolitionist is that the revolutionary cognitivist can say all of this too, *and more besides*. The revolutionary cognitivist can say all of this

---

<sup>529</sup> Garner (2010: 217).

because she agrees that reasoning about the best means to ends is possible and that these ends may very well be that of improving the human condition in situations where questions of euthanasia and abortion are at stake. And the revolutionary cognitivist too isn't constrained by the practice of *moralizing* one's thoughts and action. Moreover, the revolutionary cognitivist has further benefits that the abolitionist doesn't have. I have discussed these in the previous section. The abolitionist who simply proposes to do away with moral thought and talk and offers no replacing discourse at all cannot avail herself of these and other benefits. Revolutionary cognitivism is therefore better than abolitionism.

#### 8.4.2 Fictionalism

I have already introduced the two main forms of fictionalism above; viz., attitude- and content-fictionalism. I argue that both suffer from the same flaw .

Consider the attitude fictionalist, according to whom we should make-believe that stealing is wrong (rather than believe that stealing is wrong or believe that 'in the fiction, stealing is wrong'). To see the problem as it applies to her theory, consider the claim that the attitude of make-belief is a highly overridable attitude.<sup>530</sup> Lots of things can happen that can knock us out of the mode of pretence. A child pretends a chair is a boat, but if dinner is served she simply leaps into the 'water' and runs to the table. We pretend to know that our lover isn't cheating on us even in the face of countervailing evidence; but there comes a point where we can no longer hold up the pretence.

The problem with this aspect of fictionalism is that it sits ill with its official *raison d'être*—viz., ensuring in the best way we can that we'll live in a society that doesn't contain killers and stealers and is generally conducive to human cooperation and flourishing *without violating the truth-norm (i.e., without deluding or deceiving ourselves)*. The fictionalist thinks that we can attain this state of affairs by *pretending* that stealing is morally wrong. When it actually matters we'll have already managed to get ourselves in the mode of pretending that stealing is wrong and as a result refrain from stealing.<sup>531</sup> But if the attitude of pretence is highly overridable—for instance by the thought that stealing will be *fun* or *prudentially speaking the thing to do* (because the agent stands a lot to gain without significant

---

<sup>530</sup> Oddie and Demetriou (2010: 200).

<sup>531</sup> Miller (2013: 120).

losses)—then we can't seem to be certain about our bulwark against stealing. Unless we are involved in self-delusion such as may happen when we drop the disposition to believe error theory in contexts of heightened epistemic standards. Presumably, if one never dismisses morality then one also won't dismiss morality when it practically matters. However, by accepting that we can deceive ourselves in this way, revolutionary fictionalists abandon the truth-norm, which is not something that fictionalists accept. After all, if we can simply delude ourselves, whence the need to introduce complicating fictionalist elements into one's replacement moral discourse? So attitude-fictionalism can't work even on its own terms.

According to the content-fictionalist we should believe that 'in the fiction, stealing is wrong'. Would accepting *content*-fictionalism help? It wouldn't. For such a view allows people to step out of the moral fiction all too easily whenever they need to if they can gain a lot and stand little to lose. After all, on content-fictionalist views agents know all too well that, really, stealing isn't wrong (it is only wrong in the fiction). Again the problem is that these features of both kinds of fictionalist views sit ill with fictionalism's purpose; viz., the securing of the alleged non-moral benefits of peaceful co-existence short of deluding oneself.

Proponents of fictionalism can respond to my objections by bringing in the argument that we can *compartmentalize*, in our heads, first-order and second-order moral thinking such that our second-order nihilistic beliefs won't interfere with our first-order (fictional) moral beliefs in practical deliberation.<sup>532</sup> Compartmentalization is useful because if it works then the abovementioned tension between the purpose of fictionalism and its expected success will disappear. After all, if we can block metaethical considerations from arising when it practically matters then it *will* be possible that agents who make-believe moral propositions or believe fictionalist propositions will be set steadfastly on track to not stealing. This will allow such agents to reap the benefits of social cooperation without deluding themselves.

Let's be clear on what compartmentalization involves. Compartmentalization can be achieved through *precommitment*.<sup>533</sup> To be precommitted to morality is to choose to develop a stable disposition to think in moral terms—for instance, in the shops, when one is tempted to steal—and yet to remain disposed to 'step out' of the

---

<sup>532</sup> Olson (2011b: 199); West (2010: 195); Railton (2003: 297).

<sup>533</sup> Joyce (2005: 303-8).



moral fiction in critical contexts such as the philosophy classroom. On this model, fictional thinking doesn't feel any different from moral thinking. One isn't supposed to—and if one has properly compartmentalized one thoughts, one won't—actively rehearse the truth of the error theory and the conscious decision to enter the fiction. The temptation to steel will automatically feel abhorrent if one is precommitted to the moral fiction.

Unfortunately for the friend of fictionalism and precommitment, it seems that the theory of precommitment only tells the fictionalist story as we already had it in a slightly different vocabulary. I surmise that it is reasonable to be sceptical about whether having precommitted oneself to the moral fiction—whether this happened via attitudes or fictional operators on sentences is a moot point in this context—will reliably enough yield the result that a stable social world will be attained. It would seem to be a safe empirical conjecture that for people with different ends, such as that of living in a world full of excitement and very few rules, precommitment to morality either wouldn't work when it matters or even that these people won't try to precommit themselves to begin with. Unless of course agents *delude* themselves into thinking that morality is in order and that error theory is false. But that is something that fictionalists can't accept. It violates the truth-norm that they want to uphold, and if agents completely delude themselves into thinking that morality is real then this obfuscates the need for a fictionalist treatment of moral thought and talk. Therefore, again we have a position that fails by its own lights. As my own revolutionary cognitivism allows people to have whatever fundamental ends they want and only allows for critique of behaviour that is hypothetically normative, revolutionary cognitivism is not affected by the failing of the precommitment story.<sup>534</sup> Since revolutionary cognitivism doesn't fail and has some benefits, and since fictionalism can't work, the former is to be preferred over the latter.

### 8.4.3 Preservationism

According to preservationism, we ought to preserve moral discourse. We ought to continue to moralize without changing anything about our moral practice or moral discourse. We do have the background knowledge that error theory is true, but we simply ignore this when it matters (or we compartmentalize our metaethical and

---

<sup>534</sup> For further arguments against fictionalism see Cuneo and Christy (2011: 98); Miller (2013: Chapter 6); Lillehammer (2004); Sainsbury (2010).

ethical thoughts). The idea is that by preserving moral discourse we can continue to reap its benefits, which include having a discourse with which we can hope to enjoy the benefits of social cooperation, without putting forward a hopeless theory like fictionalism.

Preservatism comes in two kinds. There is universal preservatism, according to which all of us should continue to moralize, regardless of whether we belong to the in-crowd of philosophers acquainted with metaethics (and the true metaethical theory, the moral error theory) or to the folk who have neither the time nor the resources to appreciate the reasons for believing error theory.<sup>535</sup> And there is propagandism, according to which the error theorists should keep the truth of the error theory to themselves and pretend to be moralizing when they speak to the non-cognoscenti (out of fear that when the folk will get to know the truth they will start misbehaving—i.e., that they will start killing and stealing).<sup>536</sup> The propagandist herself, like the revolutionary cognitivist, can, when she talks to the folk, avail herself of the perlocutionary effects of presenting moral propositions, in the form of uttered statements, to an audience. In that way she can explain the use of uttering moral statements expressing moral propositions she doesn't believe in.<sup>537</sup> I object to each version of preservatism in turn.

Joyce complains about universal preservatism in the following way:

The option of carrying on as if nothing has changed—of continuing to assert moral propositions and to hold moral beliefs *even while maintaining moral error theoretic commitments*—is surely a nonstarter, for the kind of doxastic schizophrenia involved in such a life not only violates epistemic norms but can also be expected to lead to various kinds of pragmatic handicap<sup>538</sup>

I agree with Joyce that preservatists violate epistemic norms in the rationality-involving sense, such as the norm, defended in Chapter 6, only to believe what one has sufficient evidence to believe. Therefore, universal preservatism ought to be rejected. Revolutionary cognitivism doesn't face this difficulty. It is committed to respecting norms in the rationality-involving sense, including epistemic ones, and has various benefits on offer.

---

<sup>535</sup> Defended by Olson (2011b: 183).

<sup>536</sup> Defended by Cuneo and Christy (2011).

<sup>537</sup> Cuneo and Christy (2011: 95).

<sup>538</sup> Joyce (2007b: 68).

Propagandist preservatism faces the same problem, albeit on a smaller scale. Rather than being committed to advising *all* agents to violate categorical epistemic norms in the rationality-involving sense, it only advises metaethicists to do this. Nonetheless, this is only a difference in degree; not in kind. And again revolutionary cognitivism doesn't face this difficulty because it is committed to respecting norms in the rationality-involving sense, including epistemic ones and because it has various benefits on offer.

Overall, I conclude that revolutionary cognitivism wins all of the pair-wise comparisons with the different views about what we should do with moral discourse after error theory. Therefore, after error theory we should accept it as our best theory about what we prudentially ought to do with our flawed moral discourse.

## **8.5 Conclusion**

In this chapter I have argued that my version of revolutionary cognitivism is plausible because it has various benefits (§8.2). I have also argued that it can respond to various problems (§8.3) and that it wins in every pair-wise comparison of it with the other proposals abolitionism, fictionalism, and preservatism (§8.4). Combining the results of §8.2-4 yields evidence that revolutionary cognitivism should be taken seriously as a response to the truth of the kind of error theory that I have defended in Chapters 1-7. Indeed, I have argued that error theorists have a working theory about what to do after error theory, so they can also solve the fourth of the four major problems for error theory that I mentioned in §1.4. This was the problem that although error theory may be true of our actual moral discourse, it is still a highly unsatisfactory theory because it can't tell us how we should respond to its message and therefore that it can't tell us what we ought to do with our error-riddled moral discourse.

Overall, then, the kind of error theory that I have defended in this dissertation is more plausible than it is often assumed to be. Not only is error theory independently credible, it can also be combined with a view about the future of moral discourse that has various benefits, including that we can continue to make sense of our belief that most people most of the time are under genuine, rationality-involving normative force not to steal or kill.

## Bibliography

Altham, J.E.J., (1986), 'The Legacy of Emotivism', in: MacDonald, G., C. Wright (eds.), *Fact, Science and Morality: Essays on A.J. Ayer's Language, Truth and Logic*, Oxford: Basil Blackwell

Anscombe, G.E.M., (1958), 'Modern Moral Philosophy', in: *Philosophy* **33**:1-19

Anscombe, G.E.M., (1959), *An Introduction to Wittgenstein's Tractatus*, Philadelphia: University of Pennsylvania Press

Aristotle, *Nicomachean Ethics*

Armstrong, D.M., (1978), *Nominalism & Realism: Universals & Scientific Realism, Vol. 1*, Cambridge: Cambridge University Press

Armstrong, D.M., (1989), *Universals*, Boulder: Westview

Armstrong, D.M., (2006), 'The scope and limits of human knowledge', in: *Australasian Journal of Philosophy* **84**: 159-166

Austin, J.L., (1975), *How to do things with words*, Oxford: Oxford University Press

Ayer, J.A., (1936), *Language, Truth and Logic*, London: Victor Gollancz

Baillie, J., (2000), *Hume on Morality*, London : Routledge

Balaguer, M., (2011), 'Bare Bones Moral Realism and the Objections from Relativism', in: Hales, S.D. (ed.), *A Companion to Relativism*, Oxford: Wiley-Blackwell

Barnes, R.E., (2012), 'Book Review: Southwood, Nicholas, *Contractualism and the Foundations of Morality*', in: *Australasian Journal of Philosophy* **90**: 815-8

Beaver, D.I., B. Geurts, (2011), 'Presupposition', in: Zalta, E.N. (ed.), *Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/sum2011/entries/presupposition/> Visited on 12 April 2011

Bedke, M.S., (2010), 'Might All Normativity be Queer?' in: *Australasian Journal of Philosophy* **88**: 41-58

Bedke, M.S., (2012), 'Against Normative Naturalism', in: *Australasian Journal of Philosophy* **90**: 111-29

Bentham, J., (1789), *An Introduction to the Principles of Morals and Legislation*, Oxford: Clarendon Press (1907)

Biehl, J., (2008), 'The Insignificance of Choice', in: Chan, D., (ed.), (2008), *Moral Psychology Today*, Dordrecht, Springer Publishers

Blackburn, S., (1971), 'Moral Realism', in: Casey, J., (ed.), *Morality and Moral Reasoning*, London: Methuen

Blackburn, S., (1984), *Spreading the Word*, Oxford: Oxford University Press

Blackburn, S., (1985), 'Errors and the Phenomenology of Value', in: Honderich, T. (ed.), *Morality and Objectivity: A Tribute to J.L. Mackie*, London: Routledge & Kegan Paul

Blackburn, S., (1992), 'Gibbard on Normative Logic', in: *Philosophy and Phenomenological Research* **52**: 947-52

Blackburn, S., (1993), *Essays in Quasi-Realism*, Oxford: Oxford University Press

- Blackburn, S., (2005), 'Quasi-Realism no Fictionalism', in: Kalderon, M.E. (ed.), *Fictionalism in Metaphysics*, Oxford: Oxford University Press
- Bloomfield, P., (2001), *Moral Reality*, Oxford: Oxford University Press
- Bloomfield, P., (2007), 'Book Review: *The Evolution of Morality* by Richard Joyce', in: *Mind* **116**: 176-80
- Boghossian, P., (1990), 'The Status of Content', in: *The Philosophical Review* **99**: 157-84
- Boghossian, P., (2006), 'What is Relativism?' in: Greenough, P., M.P. Lynch (eds.), *Truth and Realism*, Oxford: Oxford University Press
- Bond, E.J., (1966), 'Moral Requirements and the Need for Deontic Language', in: *Philosophy* **41**: 233-57
- Boyd, R., (1988), 'How to be a Moral Realist', in: Sayre-McCord, G. (ed.), *Essays on Moral Realism*, Ithaca, NY: Cornell University Press
- Brandt, R., (1979), *A Theory of the Good and the Right*, Oxford: Clarendon Press
- Bratman, M.E., (1986), *Intentions, Plans, and Practical Reasoning*, Oxford: Oxford University Press
- Bratman, M.E., (2009), 'Intention, Belief and Instrumental Rationality', in: Sobel, D., S. Wall, (eds.), *Reasons for Action*, Cambridge: Cambridge University Press
- Brink, D.O., (1984), 'Moral Realism and the Skeptical Arguments from Disagreement and Queerness', in: *Australian Journal of Philosophy* **62**: 111-25
- Brink, D.O., (1986), 'Externalist Moral Realism', in: *Southern Journal of Philosophy* **24**: 23-41
- Brink, D.O., (1989), *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press
- Brink, D.O., (1992), 'A Puzzle About the Rational Authority of Morality', in: *Philosophical Perspectives* **6**: 1-26
- Brink, D.O., (1997), 'Kantian Rationalism: Inescapability, Authority, and Supremacy', in: Cullity, G., B. Gaut (eds.), *Ethics and Practical Reason*, Oxford: Oxford University Press
- Broome, J., (1999), 'Normative Requirements', in: *Ratio* **12**: 398-419
- Broome, J., (2007), 'Is Rationality Normative?' in: *Disputatio* **23**: 161-78
- Broome, J., (2008), 'Reply to Southwood, Kearns and Star, and Cullity', in: *Ethics* **119**: 96-108
- Brown, P., (2013), 'The impossibility of morality', in: *Philosophical Studies* **163**: 627-36
- Brunero, J., (2005) 'Two Approaches to Instrumental Rationality and Belief Consistency', in: *Journal of Ethics & Social Philosophy* **1**: 1-20
- Brunero, J., (2010), 'The Scope of Rational Requirements', in: *The Philosophical Quarterly* **60**: 28-49
- Burge, T., (1992), 'Philosophy of Language and Mind: 1950-90', in: *Philosophical Review* **101**: 3-51
- Burgess, J.A., (1998), 'Error Theories and Values', in: *Australasian Journal of Philosophy* **76**: 534-52
- Burgess, J.A., (2010), 'Against Ethics', in: Joyce, R., S. Kirchin (eds.), *A World Without Values*, Dordrecht: Springer
- Carroll, L., (1895), 'What the Tortoise Said to Achilles', in: *Mind* **4**: 278-80

Casebeer, W., (2003), *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition*, Cambridge MA: MIT Press

Clark, P., (2009), 'Mackie's Motivational Argument', in: Sobel, D., S. Wall (eds.), *Reasons for Action*, Cambridge: Cambridge University Press

Coons, C., (2011), 'How to prove that some acts are wrong (without using substantive moral premises)', in: *Philosophical Studies* **155**: 83-98

Copp, D., (1995), *Morality, Normativity and Society*, Oxford: Oxford University Press

Copp, D., (2007), *Morality in a Natural World: Selected Essays in Metaethics*, Cambridge: Cambridge University Press

Copp, D., (2009a), 'Toward A Pluralist and Teleological Theory on Normativity', in: *Philosophical Issues* **19**: 21-37

Copp, D., (2009b), 'Realist-expressivism and Conventional Implicature', in: Shafer-Landau, R., (ed.), *Oxford Studies in Metaethics* **4**: 167-202

Copp, D., (2010), 'Normativity, Deliberation, and Queerness', in: Joyce, R., S. Kirchin (eds.), *A World Without Values: essays on John Mackie's Moral Error Theory*, Dordrecht: Springer

Copp, D., (2011), 'Normativity', in: *Australasian Journal of Philosophy* **89**: 180-3

Copp, D., (2012), 'Normativity and reasons: five arguments from Parfit against normative naturalism', in: Nuccetelli, S., G. Seay (eds.), *Ethical Naturalism: Current Debates*, Cambridge: Cambridge University Press

Cuneo, T., (2007a), *The Normative Web: an argument for moral realism*, Oxford: Oxford University Press

Cuneo, T., (2007b), 'Recent Faces of Moral Nonnaturalism', in: *Philosophy Compass* **2**: 850-79

Cuneo, T., (2012), 'Moral naturalism and categorical reasons', in: Nuccetelli, S., G. Seay (eds.), *Ethical Naturalism: Current Debates*, Cambridge: Cambridge University Press

Cuneo, T., S. Christy, (2011), 'The Myth of Moral Fictionalism', in: Brady, M. (ed.), *New Waves in Metaethics*, Basingstoke: Palgrave Macmillan

Daly, C., D. Liggins, (2010), 'In defence of error theory', in: *Philosophical Studies* **149**: 209-30

Darwall, S., (1983), *Impartial Reason*, Ithaca NY: Cornell University Press

Darwall, S., (2001), 'Because I want it', in: *Social Philosophy and Policy* **18**: 129-53

Darwall, S., (2002), *Contractarianism/Contractualism*, Oxford: Blackwell

Darwall, S., (2006), 'Morality and Practical Reason: A Kantian Approach', in: Copp, D. (ed.), *The Oxford Handbook of Ethical Theory*, Oxford: Oxford University Press

Darwall, S., A. Gibbard, P. Railton, (1992), 'Toward Fin de siècle Ethics: Some Trends', in: *Philosophical Review* **101**: 115-89

Davidson, D., (2004), *Problems of Rationality*, Oxford: Oxford University Press

Davis, W., (2010), 'Implicature', in: Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/fall2012/entries/implicature/>

Denham, A.E., (2000), *Metaphor and Moral Experience*, Oxford: Oxford University Press

Doris, J., (2002), *Lack of Character: Personality and Moral Behaviour*, New York: Cambridge University Press

- Dreier, J. (1997), 'Humean Doubts about the Practical Justification of Morality', in: Cullity, G., B. Gaut (eds.), *Ethics and Practical Reason*, Oxford: Oxford University Press
- Dreier, J. (2006), 'Moral Relativism and Moral Nihilism', in: Copp, D. (ed.), *The Oxford Handbook of Ethical Theory*, 240-64
- Dreier, J., (2010), 'Mackie's Realism: Queer Pigs and the Web of Belief', in: Joyce, R., S. Kirchin (eds.), *A World Without Values, Essays on John Mackie's Moral Error Theory*, Dordrecht, the Netherlands: Springer
- Dunaway, B., A. Edmonds, D. Manley (forthcoming), 'The Folk Probably Think What you Think They Think', in: *Australasian Journal of Philosophy*
- Dworkin, R., (1996), 'Objectivity and Truth: You'd Better Believe It', in: *Philosophy and Public Affairs* **25**: 87-139
- Egan, A., (2007), 'Quasi-Realism and Fundamental Moral Error', in: *Australasian Journal of Philosophy* **85**: 205-19
- Ellis, J., (2006), 'Color, Error and Explanatory Power', in: *Dialectica* **60**: 171-79
- Enoch, D., (2005), 'Why Idealize?' in: *Ethics* **115**: 759-87
- Enoch, D., (2011a), *Taking Morality Seriously: A Defence of Robust Realism*, Oxford: Oxford University Press
- Enoch, D., (2011b), 'Schmagency, Revisited', in: Brady, M. (ed.), *New Waves in Metaethics*, Basingstoke: Palgrave Macmillan
- Evans, M., Shah, N., (2012), 'Mental Agency and Metaethics', in: Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics* **7**: 81-109
- Farber, P.L., (1998), *The Temptations of Evolutionary Ethics*, Berkely CA: University of California Press
- Farrell, R., (1981), 'Metaphysical Necessity is not Logical Necessity', in: *Philosophical Studies* **39**: 141-53
- Fehige, C., (2001), 'Instrumentalism', in: Millgram, E. (ed.), *Varieties of Practical Reasoning*, Massachusetts, MA: MIT Press
- Field, H., (1973), 'Theory Change and the Indeterminacy of Reference', in: *Journal of Philosophy* **70**: 462-81
- Field, H., (1989), *Realism, mathematics and modality*, Oxford: Blackwell
- Fine, K., (2002), 'The Varieties of Necessity', in: Gendler, T.S., Hawthorne, J., (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press
- Fine, K., (2005), 'Necessity and Non-Existence', in: Fine, K., (2005), *Modality and Tense*, Oxford: Oxford University Press
- Finlay, S., (2005), 'Value and Implicature', in: *Philosopher's Imprint* **5**: 1-20
- Finlay, S., (2008a), 'The Error in the Error Theory', in: *Australasian Journal of Philosophy* **86**: 347 – 369
- Finlay, S., (2008b), 'Too Much Morality', in: Bloomfield, P. (ed.), *Morality and Self-Interest*, Oxford: Oxford University Press
- Finlay, S., (2009a), 'Against All Reason? Scepticism about the Instrumental Norm', in: Pigden, C.R., (ed.), *Hume on Motivation and Virtue*, London: Palgrave Macmillan
- Finlay, S., (2009b), 'The Obscurity of Internal Reasons', in: *Philosophers' Imprint* **9**: 1-22

- Finlay, S., (2011), 'Errors upon Errors: A Reply to Joyce', in: *Australasian Journal of Philosophy* **89**: 535-47
- Fischer, J., M. Ravizza, (1998), *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press
- Fitzpatrick, W.J., (2008), 'Robust Ethical Realism, Non-Naturalism and Normativity', in: *Oxford Studies in Metaethics* **3**, Oxford: Oxford University Press
- Foot, P., (1972), 'Morality as a System of Hypothetical Imperatives', in: *Philosophical Review* **81**: 305-16
- Frankena, W., (1958), 'Obligation and Motivation in Recent Moral Philosophy', in: Melden, A.I. (ed.), *Essays in Moral Philosophy*, Seattle: University of Washington Press
- Freeman, S., (2012), 'Original Position', in: Zalta, E.N. (ed.), *Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/spr2012/entries/original-position/>
- Gampel, E.H., (1996), 'A Defense of the Autonomy of Ethics: Why Value Is Not Like Water', in: *Canadian Journal of Philosophy* **26**: 191-209
- Garner, R., (1990), 'On the Genuine Queerness of Moral Properties and Facts', in: *Australasian Journal of Philosophy* **68**: 137-46
- Garner, R., (1994), *Beyond Morality*, Philadelphia: Temple University Press
- Garner, R., (2010), 'Abolishing Morality', in: Joyce, R., S. Kirchin (eds.), *A World Without Values: Essays on John Mackie's Moral Error Theory*, Dordrecht: Springer
- Gauthier, D., (1986), *Morals by Agreement*, Oxford: Clarendon Press
- Gert, J., (2004), *Brute Rationality: Normativity and Human Action*, Cambridge: Cambridge University Press
- Gibbard, A., (1990), *Wise Choices, Apt Feelings*, Cambridge Massachusetts, Harvard University Press
- Gibbard, A., (2003), *Thinking How to Live*, Cambridge MA: Harvard University Press
- Gill, M., (1996), 'A philosopher in his closet: Reflexivity and justification in Hume's moral theory', in: *Canadian Journal of Philosophy* **26**: 231-56
- Gill, M.B., (2008), 'Variability and moral phenomenology', in: *Phenomenology and Cognitive Science* **7**: 99-113
- Glassen, P., (1959), 'The Cognitivity of Moral Judgments', in: *Mind* **68**: 57-72
- Greco, J., (2007), *Putting Skeptics in their Place : the nature of skeptical arguments and their role in philosophical inquiry*, Cambridge: Cambridge University Press
- Green, M., (2009), 'Speech Acts', in: Zalta, E.N., (ed.), *Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/spr2009/entries/speech-acts/>.
- Grice, H.P., (1975), 'Logic and Conversation', in: Cole, P., J. Morgan (eds.), *Syntax and Semantics, 3: Speech Acts*, New York: Academic Press
- Grice, H.P., (1989), *Studies in the Ways of Words*, Cambridge MA: Harvard University Press



- Hägerström, A., (1953), *Inquiries into the Nature of Law and Morals*, Stockholm: Almqvist & Wiksell
- Haji, I., (1998), *Moral Appraisability*, Oxford: Oxford University Press
- Haji, I., (2003), *Deontic Morality and Control*, Cambridge: Cambridge University Press
- Hampton, J., (1998), *The Authority of Reason*, Cambridge: Cambridge University Press
- Hare, R.M., (1981), *Moral Thinking: Its Levels, Point and Method*, Oxford: Oxford University Press
- Hare, R.M., (1999), *Objective Prescriptions and Other Essays*, Oxford: Clarendon Press
- Harman, G., (1975), 'Moral Relativism Defended', in: *Philosophical Review* **84**: 3-22
- Harman, G., (1984), 'Is There a Single True Morality?' in: Copp, D., Zimmerman, D., (eds.), (1984), *Morality, Reason and Truth – New Essays on the Foundations of Ethics*, Totowa: Rowman & Allanheld
- Harman, G., (1986), *Change in View*, Cambridge MA: MIT Press
- Harman, G., J.J. Thomson, (1994), *Moral Relativism and Moral Objectivity*, Oxford: Blackwell
- Hartcourt, E., (2005), 'Quasi-Realism and Ethical Appearances', in: *Mind* **114**: 249-75
- Heathwood, C., (2011), 'Desire-Based Theories of Reasons, Pleasure and Welfare', in: Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics* **6**
- Heathwood, C., (2012), 'Could Morality Have a Source?' In: *Journal of Ethics and Social Philosophy* **6**: 1-20
- Heumer, M., (2005), *Ethical Intuitionism*, Basingstoke: Palgrave Macmillan
- Hinckfuss, I., (1987), *The moral society: Its structure and effects*, Canberra: Australian National University
- Horgan, T., M. Timmons, (1992a), 'Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived', in: *Philosophical Papers* **21**: 153-75
- Horgan, T., M. Timmons, (1992b), 'Troubles on Moral Twin Earth: Moral Queerness Revived', in: *Synthese* **92**: 221-60
- Horgan, T., M. Timmons, (2002), 'Conceptual Relativity and Metaphysical Realism', in: *Philosophical Issues* **12**: 74-96
- Horgan, T., M. Timmons, (2005), 'Moral Phenomenology and Moral Theory', in: *Philosophical Issues* **15**: 56-77
- Horgan, T., M. Timmons, (2008a), 'Prolegomena to a future phenomenology of morals', in: *Phenomenology and the Cognitive Sciences* **7**: 115-31
- Horgan, T., M. Timmons, (2008b), 'What Does Moral Phenomenology Tell Us About Moral Objectivity?', in: *Social Philosophy and Policy* **25**: 267-300
- Hubin, D., (1999), 'What's Special About Humeanism', in: *Nous* **33**: 30-45
- Hubin, D., (2001), 'The Groundless Normativity of Instrumental Rationality', in: *The Journal of Philosophy* **98**: 445-468
- Huemer, M., (2005), *Ethical Intuitionism*, Basingstoke: Palgrave MacMillan
- Hume, D., (1739-40) [1969], *A Treatise of Human Nature*, Harmondsworth: Penguin Books
- Hume, D., (1751), *Enquires Concerning Human Understanding and Concerning the Principles of Morals*, (ed. Selby-Bigge), Oxford: Clarendon Press

Hussain, N.J.Z., (2010), 'Error Theory and Fictionalism', in: Skorupski, J., (ed.), (2010), *The Routledge Companion to Ethics*, Abingdon: Routledge

Jackson, F., (1998), *From Metaphysics to Ethics*, Cambridge: Cambridge University Press

Jackson, F., (2012), 'On ethical naturalism and the philosophy of language', in: Nuccetelli, S., G. Seay (eds.), *Ethical Naturalism: Current Debates*, Cambridge: Cambridge University Press

Jackson, F., P. Pettit, (1995), 'Moral Functionalism and Moral Motivation', in: *The Philosophical Quarterly* **45**: 20-40

James, A., (2007), 'Constructivism about Practical Reasons', in: *Philosophy and Phenomenological Research* **74**: 302-25

Johnston, M., (1987), 'Human Beings', in: *The Journal of Philosophy* **84**: 59-83

Johnston, M., (2010), *Surviving Death*, Princeton NJ: Princeton University Press

Joyce, R., (2000), 'The Fugitive Thought', in: *Journal of Value Inquiry* **34**: 463-78

Joyce, R., (2001), *The Myth of Morality*, Cambridge: Cambridge University Press

Joyce, R., (2005), 'Moral Fictionalism', in: Kalderon, M.E. (ed.), *Fictionalism in Metaphysics*, Oxford: Oxford University Press

Joyce, R., (2006), *The Evolution of Morality*, Cambridge MA: The MIT Press

Joyce, R., (2007a), 'Moral Anti-Realism', in: Zalta, E.N., (ed.), *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/fall2008/entries/moral-anti-realism/>. Visited on 31 March 2010

Joyce, R., (2007b), 'Morality, Schmorality', in: Bloomfield, P., (ed.), (2007), *Morality and Self-Interest*, Oxford: Oxford University Press

Joyce, R., (2009), 'Is Projectivism Empirically Tractable?' in: *Ethical Theory and Moral Practice* **12**: 53-75

Joyce, R., (2010a), 'Patterns of Objectification', in: Joyce, R., S. Kirchin (eds.), *A World Without Values: essays on John Mackie's Moral Error Theory*, Dordrecht: Springer

Joyce, R., (2011a), 'The Error in "The Error in the Error Theory"', in: *Australasian Journal of Philosophy* **89**: 519-34

Joyce, R., (2011b), 'The Accidental Error Theorist', in: Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics* **6**: 153-180

Joyce, R., (2012), 'Metaethical pluralism: How both moral naturalism and moral skepticism may be permissible positions', in: Nuccetelli, S., G. Seay (eds.), *Ethical Naturalism: Current Debates*, Cambridge: Cambridge University Press

Joyce, R., (2013), 'Review of David Enoch's *Taking Morality Seriously*', in: *Ethics* **123**: 365-9

Joyce, R., (MS), 'Enough with the errors! A final reply to Finlay' (available at [http://www.victoria.ac.nz/staff/richard\\_joyce/onlinepapers.html](http://www.victoria.ac.nz/staff/richard_joyce/onlinepapers.html))

Juhl, C., E. Loomis, (2010), *Analyticity*, London: Routledge

Kagan, S., (1989), *The Limits of Morality*, Oxford: Oxford University Press

Kahane, G., (2013), 'Must Metaethical Realism Make a Semantic Claim?' in: *Journal of Moral Philosophy* **10**: 148-78

Kalderon, M.E., (2005), *Moral Fictionalism*, Oxford: Oxford University Press

Kalderon, M.E., (2008), 'Moral Fictionalism: Summary', in: *Philosophical Books* **49**: 1-3

Kalf, W.F., (2013), 'Moral Error Theory, Entailment and Presupposition', in: *Ethical Theory and Moral Practice* **16**: 923-37

Kalf, W.F., (MS), 'Is Morality Impossible?'

- Kane, R., (1996), *The Significance of Free Will*, Oxford: Oxford University Press
- Karttunen, L., (1973), 'Presuppositions of Compound Sentences', in: *Linguistic Inquiry* **4**: 167-93
- Kavka, G.S., (1984), 'A Reconciliation Project', in: Copp, D., D. Zimmerman (eds), *Morality, Truth and Reason*, Totowa, NJ: Rowman and Allanheld
- Kiesewetter, B., (2012), 'A dilemma for Parfit's conception of normativity', in: *Analysis* **72**: 466-74
- Kirchin, S., (2003), 'Ethical Phenomenology and Metaethics', in: *Ethical Theory and Moral Practice* **6**: 241-64
- Kirchin, S., (2010), 'A Tension in the Moral Error Theory', in: Joyce, R., S. Kirchin (eds.), *A World Without Values: Essays on John Mackie's Moral Error Theory*, Dordrecht, the Netherlands: Springer
- Kolodny, N., (2005), 'Why be Rational?' in: *Mind* **114**: 509-64
- Kolodny, N., (2008), 'Why Be Disposed to Be Coherent?', in: *Ethics* **118**: 437-63
- Kornblith, H., (2012), *On Reflection*, Oxford: Oxford University Press
- Korsgaard, C., (1986), 'Skepticism about Practical Reason', in: *The Journal of Philosophy* **83**: 5-25
- Korsgaard, C., (1996), *The Sources of Normativity*, Cambridge: Cambridge University Press
- Korsgaard, C., (1997), 'The Normativity of Instrumental Reason', in: Cullity, B., B. Gaut (eds.), (1997), *Ethics and Practical Reason*, Oxford: Clarendon Press
- Kripke, S.A., (1972), *Naming and Necessity*, Oxford: Basil Blackwell
- Lau, J., M. Deutsch (2002), 'Externalism About Mental Content', in: *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/content-externalism/>
- Lewis, D.K., (1978), 'Truth in Fiction', in: *American Philosophical Quarterly* **15**: 37-46
- Lewis, D.K., (1986), *On the Plurality of Worlds*, Oxford: Basil Blackwell
- Lewis, D.K., (1989), 'Dispositional Theories of Value', in: *Proceedings of the Aristotelian Society, Suppl. Vol.* **63**: 113-37
- Lewis, D.K., (1996), 'Desire as Belief II', in: *Mind* **105**: 303-313
- Libet, B., (2004), *Mind Time: The Temporal Factor in Consciousness*, Cambridge MA: Harvard University Press
- Lillehammer, H., (1999), 'Normative Antirealism', in: *The Southern Journal of Philosophy* **37**: 201-25
- Lillehammer, H., (2000), 'Revisionary Dispositionalism and Practical Reason', in: *The Journal of Ethics* **4**: 173-90
- Lillehammer, H., (2004), 'Moral Error Theory', in: *Proceedings of the Aristotelian Society* **104**: 95-111
- Lillehammer, H., (2007), *Companions in Guilt: Arguments for Ethical Objectivity*, New York: Palgrave
- Loeb, D., (2008), 'Moral incoherentism: How to pull a metaphysical rabbit out of a semantic hat', in: Sinnott-Armstrong, W. (ed.), *Moral Psychology*, Vol. 2, Cambridge MA: MIT Press

- Loeb, D., (2010), 'The Argument from Moral Experience', in: Joyce, R., S. Kirchin (eds.), *A World Without Values: essays on John Mackie's Moral Error Theory*, Dordrecht: Springer
- Lukes, S., (1985), 'Taking Morality Seriously', in: Honderich, T. (ed.), *Morality and Objectivity*, London: Routledge
- Lycan, W.G., (1999), *Philosophy of Language: a contemporary introduction*, London: Routledge
- MacIntyre, A., (1981), *After Virtue*, London: Duckworth
- Mackie, J.L., (1946), 'A Refutation of Morals', in: *Australasian Journal of Philosophy* **24**: 77– 90
- Mackie, J.L., (1976), *Problems from Locke*, Oxford: Clarendon Press
- Mackie, J.L., (1977), *Ethics: inventing right and wrong*, London: Penguin Publishers
- Mackie, J.L., (1980), *Hume's moral theory*, London: Routledge and Kegan Paul
- Mandelbaum, M., (1955), *The Phenomenology of Moral Experience*, Glencoe IL.: Free Press
- Markovits, J., (2011), 'Why be an Internalist about Reasons?' in: Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics* **6**
- Marks, J., (2013), *Ethics without Morals: In Defense of Amoralism*, London: Routledge
- McDowell, J., (1985), 'Values and Secondary Qualities', in: Honderich, T. (ed.), *Morality and Objectivity: Essays in Memory of J.L. Mackie*, London: Routledge & Kegan Paul
- McGrath, M., (2004), 'Moral Knowledge by Perception', in: *Philosophical Perspectives* **18**: 219-28
- McGrath, M., (2012), 'Propositions', in: Zalta, E.N. (ed.), *Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/entries/propositions/>, visited on 12 September 2012
- McLaughlin, B., K. Bennett, (2010), 'Supervenience', in: Zalta, E.N., (ed.), *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/sum2010/entries/supervenience/> Visited on May 25 2010
- McNaughton, D., (1988), *Moral Vision*, Basil Blackwell: London.
- Merricks, T., (2001), *Objects and Persons*, Oxford: Oxford University Press
- Miller, A., (2002), 'Wright's Argument against Error Theories', in: *Analysis* **62(2)**: 98-103
- Miller, A., (2003), *An Introduction to Contemporary Metaethics*, Cambridge: Polity Press
- Miller, A., (2013), *An Introduction to Contemporary Metaethics* (2<sup>nd</sup> ed.), Cambridge: Polity Press
- Miller, C., (2009), 'The Conditions of Moral Realism', in: *The Journal of Philosophical Research* **34**: 123-155
- Moore, G.E., (1903), *Principia Ethica*, Cambridge: Cambridge University Press
- Moore, G.E., (1942), 'A Reply to My Critics', in: Schilpp, P.A., (ed), *The Philosophy of G. E. Moore*: Evanston ILL: Northwestern University Press
- Moore, G.E., (1953), *Some Main Problems of Philosophy*, London: George, Allen and Unwin

Nagel, T., (1970), *The Possibility of Altruism*, Princeton: Princeton University Press  
 Nagel, T., (1986), *The view from nowhere*, Oxford: Oxford University Press

Nesbitt, W., (1977), 'Categorical Imperatives – A Defense', in: *The Philosophical Review* **81**: 305-16

Newman, J., (1981), 'The Fictionalist Analysis of Some Moral Concepts', in: *Metaphilosophy* **12**: 47-56

Nichols, S., (2004), *Sentimental Rules: On the Natural Foundations of Moral Judgment*, Oxford: Oxford University Press

Nietzsche, F., (1883-88), *The Will to Power*, Trans. R. Hollingdale and W. Kaufmann; ed. W. Kaufmann [1986], New York: Vintage

Nolan, D., G., Restall, C. West, (2005), 'Moral fictionalism versus the rest', in: *Australasian Journal of Philosophy* **83**: 307-30

Nuccetelli, S., G. Seay (2012), 'Does analytic moral naturalism rest on a mistake?' in: Nuccetelli, S., G. Seay (eds.), *Ethical Naturalism: Current Debates*, Cambridge: Cambridge University Press

Oddie, G., D. Demetriou, (2010), 'The Fictionalist's Attitude Problem', in: Joyce, R., S. Kirchin (eds.), *A World Without Values: Essays on John Mackie's Moral Error Theory*, Dordrecht, the Netherlands: Springer

Olson, J., (2009a), 'Error Theory and Reasons for Belief', in: Reisner, A., A. Steglich-Petersen, (eds.), *Reasons for Belief*, Dordrecht, the Netherlands: Springer Publishers

Olson, J., (2009b), 'Reasons and the New Non-Naturalism', in: Robertson, S. (ed.), *Spheres of Reason: New Essays in the Philosophy of Normativity*, Oxford: Oxford University Press

Olson, J., (2011a), 'In Defense of Moral Error Theory', in: Brady, M. (ed.), *New Waves in Metaethics*, Basingstoke: Palgrave Macmillan

Olson, J., (2011b), 'Getting Real about Moral Fictionalism': in: Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics* **6**, Oxford: Oxford University Press

Olson, J., (2011c), 'Projectivism and Error in Hume's Ethics', in: *Hume Studies* **37**: 19-42

Olson, J., (2012), 'Skorupski's Middle Way in Metaethics', in: *Philosophy and Phenomenological Research* **85**: 192-200

Olson, J., (forthcoming-a), 'The Metaphysics of Reasons', in: Star, D., (ed.), *The Oxford Handbook of Reasons and Normativity*. Oxford: Oxford University Press

Olson, J., (forthcoming-b), 'Mackie's Motivational Queerness Argument Reconsidered', in: *The International Journal for the Study of Skepticism*

Olson, J., (MS), *Moral Error Theory: History, Critique, Defence*

O'Neill, O., (1988/9), 'Constructivism in Ethics', in: *Proceedings of the Aristotelian Society* **89**: 1-18

Papineau, D., (2007), 'Naturalism', in: Zalta, E.N., (ed.), *Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/entries/naturalism/> Visited on 29 July 2010

Parfit, D., (1984), *Reasons and Persons*, Oxford: Clarendon Press

Parfit, D., (1997), 'Reasons and Motivation', in: *Proceedings of the Aristotelian Society* **71**, Suppl. Vol.: 99-130

Parfit, D., (2012), *On What Matters*, Oxford: Oxford University Press (2 Volumes)

Pauer-Studer, H., (2009), 'Humean Sources of Normativity', in: Pigden, C.R. (ed.), *Hume on Motivation and Virtue*, London: Palgrave Macmillan

Pereboom, D., (2001), *Living Without Free Will*, Cambridge, Cambridge University Press

Perry, R.B., (1926), 'Value as Any Object of Interest', in his *General Theory of Value*, Cambridge MA: Harvard University Press

Philips, D., (2010), 'Mackie on Practical Reason', in: Joyce, R., S. Kirchin (eds.), *A World Without Values: essays on John Mackie's Moral Error Theory*, Dordrecht: Springer

Pigden, C.R., (2010), 'Nihilism, Nietzsche, and the Doppelganger Problem', in: Joyce, R., S. Kirchin (eds.), *Morality in a Natural World: Essays on John Mackie's Moral Error Theory*, Dordrecht, the Netherlands: Kluwer

Plato, *The Republic*

Platts, M., (1980), 'Moral Reality and the End of Desire', in: Platts (ed.), *Reference, Truth and Reality: Essays on the Philosophy of Language*, London: Routledge & Kegan Paul

Platts, M., (1991), *Moral Realities*, London: Routledge

Potts, C., (2007), 'Into the Conventional-Implicature Dimension', in: *Philosophy Compass* **2**: 665-79

Putnam, H., (1975), 'The Meaning of "meaning"', in: Putnam, H., *Mind, Language and Reality: Philosophical Papers Vol. 2*, Cambridge: Cambridge University Press

Putnam, H., (1988), *Representation and Reality*, Cambridge MA: MIT Press

Putnam, H., (1992), "Brains in a Vat", in DeRose, K., T.A. Warfield (eds.), *Skepticism: a Contemporary Reader*, Oxford: Oxford University Press,

Quine, W.V.O., (1948), 'On What There Is', in: *Review of Metaphysics* **2**: 21-38

Railton, P., (1986), 'Moral Realism', in: *Philosophical Review* **95**: 163-207

Railton, P., (1989), 'Naturalism and Prescriptivity', in: *Social Philosophy and Policy* **7**: 51-74

Railton P (1997) On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action. In: Cullity, G., B. Gaut (eds.), *Ethics and Practical Reason*, Oxford: Oxford University Press

Railton, P., (2003), *Facts, Values and Norms: Essays toward a Morality of Consequence*, Cambridge: Cambridge University Press

Railton, P., (2006), 'Humean Theory of Practical Rationality', in: Copp, D., *The Oxford Handbook of Ethical Theory*, Oxford: Oxford University Press

Railton, P., (2010), 'Realism and Its Alternatives', in: Skorupski, J., (ed.), (2010), *The Routledge Companion to Ethics*, Abingdon: Routledge

Ramsey, F., (1990), *Philosophical Papers*, Cambridge: Cambridge University Press

Rauscher, F., (2002), 'Kant's Moral Anti-Realism', in: *Journal of the History of Philosophy* **44**: 477-99

Rawls, J., (1971), *A Theory of Justice*, Cambridge MA: Harvard University Press

Raz, J., (1999), 'Explaining Normativity: On Rationality and the Justification of Reason', in: *Ratio* **12**: 354-79

Ridge, M., S. Kohler, (forthcoming), 'Revolutionary Expressivism', in: *Ratio*

- Rippon, S., (2011), 'In Defense of the Wide-Scope Instrumental Principle', In: *Journal of Ethics & Social Philosophy* **5**: 1-20
- Robertson, S., (2008), 'How to be an Error Theorist about Morality', in: *Polish Journal of Philosophy* **2**: 107-25
- Robertson, S., (2010), 'Reasons, values, and morality', in: Skorupski, J., (ed.), (2010), *The Routledge Companion to Ethics*, Abingdon: Routledge
- Robinson, D., (2009), 'Moral Functionalism, Quasi-Relativism, and the Plan', in: Braddon-Mitchell, D., R. Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*, Cambridge MA: MIT Press
- Roskies, A., (2003), 'Are Ethical Judgments Intrinsically Motivational? Lessons from Acquired Psychopathy', in: *Philosophical Psychology* **16**: 51-66
- Rowland, R., (2013), 'Moral Error Theory and the Argument from Epistemic Reasons', In: *Journal of Ethics & Social Philosophy* **7**
- Russell, B., (1922), 'Is there an absolute good?' in: Pigden, C. (ed.), (1999), *Russell on Ethics*, London: Routledge
- Russell, L., (2011), 'Two Kinds of Normativity: Korsgaard v. Hume', in: Pigden, C.R. (ed.), *Hume on Motivation and Virtue*, London: Palgrave MacMillan
- Saeed, J.I., (2003), *Semantics*, Malden MA: Blackwell Publishers
- Sainsbury, R.M., (1998), 'Projections and Relations', in: *The Monist* **81**: 130-60
- Sainsbury, R.M., (2010), *Fiction and Fictionalism*, London: Routledge
- Sayre-McCord, G., (1986), 'The Many Moral Realisms', in: *The Spindel Conference: The Southern Journal of Philosophy, Suppl. Volume* **24**: 1-22
- Scanlon, T.M., (1982), 'Contractualism and Utilitarianism', in: Sen, A., B. Williams (eds.), *Utilitarianism and Beyond*, Cambridge: Cambridge University Press
- Scanlon, T.M., (1998), *What We Owe To Each Other*: Cambridge MA: Harvard University Press
- Scanlon, T.M., (2007), 'Structural Irrationality', in: Brennan, G, R.E. Goodin, F. Jackson, M. Smith (eds.), *Common Themes: Themes from the Philosophy of Philip Pettit*, Oxford, Oxford University Press
- Scanlon, T.M., (2009), *Being Realistic about Reasons: The John Locke Lectures*, University of Oxford
- Scanlon, T.M., (2011), 'The Unity of the Normative', in: *Philosophical Studies* **154**: 443-50
- Schiffer, S., (1990), 'Meaning and Value', in: *The Journal of Philosophy* **84**: 602-14
- Schiffer, S., (2003), *The Things We Mean*, Oxford: Oxford University Press
- Schroeder, M., (2007), *Slaves of the Passions*, Oxford: Oxford University Press
- Schroeder M (2009) Means-end coherence, stringency, and subjective reasons. *Philosophical Studies* **143**: 223-48
- Schroeder, M., (2010), *Noncognitivism in Ethics*, London: Routledge
- Setiya, K., (2007), 'Cognitivism about instrumental reason', in: *Ethics* **117**: 649-73
- Shafer-Landau, R., (2003), *Moral Realism*, Oxford: Clarendon Press
- Shafer-Landau, R., (2005), 'Error Theory and the Possibility of Normative Ethics', in: *Philosophical Issues* **15**: 107-20



- Shafer-Landau, R., T. Cuneo, (2007), (eds.), *Foundations of Ethics: an Anthology*, London: Blackwell
- Shah, H., (2010), 'The Limits of Normative Detachment', In: *Proceedings of the Aristotelian Society* **110**: 347-71
- Sharp, F.C., (1928), *Ethics*, New York: The Century
- Shepski, L., (2008), 'The Vanishing Argument from Queerness', in: *The Australasian Journal of Philosophy* **86**: 371-87
- Silverstein, M., (2012), 'Inescapability and Normativity', in: *Journal of Ethics and Social Philosophy* **6**: 1-26
- Simons, M., (2006), 'Foundational Issues in Presupposition', in: *Philosophy Compass* **1**: 357-72
- Simons, M., J. Tonhauser, D. Beaver, C. Roberts (2010), 'What projects and why', in: *Proceedings of SALT* **20**: 309-27
- Singer, P., (1972), 'Famine, Affluence and Morality', in: *Philosophy and Public Affairs* **1**: 229-43
- Sinnott-Armstrong, W. (2006), *Moral Skepticisms*, New York: Oxford University Press
- Sinnott-Armstrong, W., (2008), 'Is moral phenomenology unified?', in: *Phenomenology and Cognitive Science* **7**: 85-97
- Skorupski, J., (1999), 'Irrealist Cognitivism', in: *Ratio* **12**: 436-59
- Skorupski, J., (2007), 'Internal Reasons and the Scope of Blame', in: Thomas, A. (ed.), *Bernard Williams*, Cambridge: Cambridge University Press
- Skorupski, J., (2010), *The Domain of Reasons*, Oxford: Oxford University Press
- Smith, M., (1993), 'Objectivity and Moral Realism: On the Significance of the Phenomenology of Moral Experience', in: Haldane, J., C. Wright (eds.), *Reality, Representation and Projection*, Oxford: Oxford University Press
- Smith, M., (1994), *The Moral Problem*, Oxford: Basil Blackwell
- Smith, M., (1995), 'Internal Reasons', in: *Philosophy and Phenomenological Research* **55**: 109-31
- Smith, M., (2004) Instrumental Desires, Instrumental Rationality. In: *Proceedings of the Aristotelian Society Suppl. Volume* **78**: 93-109
- Smith, M., (2006), 'Is that all there is?' in: *Journal of Ethics* **10**: 75-106
- Smith, M., (2007), 'Is There a Nexus Between Reasons and Rationality?' in: Tenenbaum, S. (ed.), *Moral Psychology*, Amsterdam: Rodopi
- Smith, M., (2009), 'Desires, Values, Reasons, and the Dualism of Practical Reason', in: *Ratio* **22**: 98-125
- Smith, M., (2010), 'Beyond the Error Theory', in: Joyce, R., S. Kirchin (eds.), *A World Without Values: essays on John Mackie's Moral Error Theory*, Dordrecht: Springer
- Smith, M., (2011), 'Deontological Moral Obligations and Non-Welfarist Agent-Relative Values', in: *RATIO* **24**: 351-63
- Smith, M., (2012a), 'Naturalism, absolutism, relativism', in: Nuccetelli, S., G. Seay (eds.), *Ethical Naturalism: Current Debates*, Cambridge: Cambridge University Press
- Smith, M., (2012b), 'A Puzzle about Internal Reasons', in: Heuer, U., G. Lang (eds.), *Luck, Value, and Commitment: Themes from the Ethics of Bernard Williams*, Oxford: Oxford University Press



Smith M (MS) A Constitutivist Theory of Reasons: Its Promise and Parts. Available at: [www.princeton.edu/~msmith/](http://www.princeton.edu/~msmith/)

Soames, S., (2006), 'Entailment, Presupposition, and Implicature', in: Borchers, D.M., (2006), *Encyclopedia of Philosophy*, Detroit MI, MacMillan

Sobel, J.H., (MS), *Good and Gold: A Judgmental History of Metaethics from Moore through Mackie*

Southwood, N., (2008), 'Vindicating the Normativity of Rationality', in: *Ethics* **119**: 9-30

Southwood, N., (2010), *Contractualism & the Foundations of Morality*, Oxford: Oxford University Press

Stalnaker, R.C., (1999), *Context and Content: Essays on Intentionality in Speech and Thought*, Oxford: Oxford University Press

Stevenson, C.L., (1937), 'The emotive meaning of ethical terms', in: *Mind* **46**: 14-31

Strandberg, C., (2012), 'A Dual Aspect Account of Moral Language', in: *Philosophy and Phenomenological Research* **84**: 87-122

Stratton-Lake, P., (2002), 'Introduction', in: Ross, W.D., (2002) [1930], *The Right and the Good*, Oxford, Oxford University Press

Strawson, P.F., (1956), 'On Referring', in: Flew, A., (ed.), *Essays in Conceptual Analysis*, London: Macmillan & Co.

Strawson, P.F., (1962), 'Freedom and Resentment', in: *Proceedings of the British Academy* **48**: 1-25

Street, S., (2009), 'In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters', in: *Philosophical Issues* **19**: 273-98

Street, S., (2010), 'What is Constructivism in Ethics and Metaethics?' in: *Philosophy Compass* **5**: 363-84

Streumer, B., (2007), 'Reasons and Entailment', in: *Erkenntnis* **66**: 353-74

Streumer, B., (2008), 'Are there irreducibly normative properties?' in: *Australasian Journal of Philosophy* **86**: 537-61

Streumer, B., (2011), 'Are normative properties descriptive properties?' in: *Philosophical Studies* **154**: 325-48

Streumer, B., (forthcoming) 'Can We Believe The Error Theory?' in: *The Journal of Philosophy*

Sturgeon, N.L., (1984), 'Moral Explanations', in: Copp, D., Zimmerman, D., (eds.), (1984), *Morality, Reason and Truth – New Essays on the Foundations of Ethics*, Totowa: Rowman & Allanheld

Sturgeon, N.L., (2006), 'Ethical Naturalism', in: Copp, D., *The Oxford Handbook of Ethical Theory*, Oxford: Oxford University Press

Suikkanen, J., (2013), 'Moral Error Theory and the Belief Problem', in: *Oxford Studies in Metaethics* **8**

Svavarsdóttir, S., (2001), 'Objective Values: Does Metaethics Rest on a Mistake?' in: Leiter, B., (ed) *Objectivity in Law and Morals*, Cambridge University Press, Cambridge

Svoboda, T., (2011), 'Hybridizing Moral Expressivism and Moral Error Theory', in: *Journal of Value Inquiry* **45**: 27-48

Thomson JJ (2008) *Normativity*. Open Court, Chicago ILL

Timmons, M., (1999), *Morality without Foundations: A Defense of Ethical Contextualism*, Oxford: Oxford University Press

Tresan, J., (2009), 'Role-Based Interpretations of Moral Judgments: An Objectivist Account', in: *Social Theory and Practice* **35**: 369-91

Tresan, J., (2010), 'Question authority: in defense of moral naturalism without clout', in: *Philosophical Studies* **150**: 221-38

Vaidya, A., (2010), 'The Epistemology of Modality', in: Zalta, E.N., (ed.), *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/entries/modality-epistemology>  
Visited on September 1 2010

Väyrynen, P., (2009), 'Normative Appeals to the Natural', in: *Philosophy and Phenomenological Research* **79**: 279-314

Velleman, J.D., (1996), 'The Possibility of Practical Reason', in: *Ethics* **106**: 694-726

Velleman, J.D., (2000), *The Possibility of Practical Reason*, Oxford: Oxford University Press

Velleman, J.,D., (2004), 'Replies to Discussion on *The Possibility of Practical Reason*', in: *Philosophical Studies* **121**: 225-38

Vogelstein, E., (forthcoming), 'Moral normativity', in: *Philosophical Studies*

Wallace, R.J., (MS), 'The Deontic Structure of Morality', available online at <http://philosophy.berkeley.edu/people/files/21>.

Way, J., (2012), 'Explaining the Instrumental Principle', in: *Australasian Journal of Philosophy* **90**: 487-506

Wayne, D., (2008), 'Implicature', in: Zalta, E.N., (ed.), *Stanford Encyclopaedia of Philosophy*, URL = <http://plato.stanford.edu/entries/implicature/#2>. Visited on April 6 2010

Wedgwood, R., (2001), 'Conceptual Role Semantics for Moral Terms', in: *Philosophical Review* **110**: 1-30

Wedgwood, R., (2007), *The Nature of Normativity*, Oxford: Oxford University Press

Wedgwood, R., (2011), 'Instrumental Rationality', in: Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics* **6**: 280-309

West, C., (2010), 'Business as Usual? The Error Theory, Internalism, and the Function of Morality', in: Joyce, R., S. Kirchin (eds.), *A World Without Values: Essays on John Mackie's Moral Error Theory*, Dordrecht, the Netherlands: Springer

Williams, B., (1979), 'Internal and external reasons', in: Harrison, R. (ed.), *Rational Action*, Cambridge: Cambridge University Press; reprinted in: Williams, B., (1981), *Moral Luck*, Cambridge: Cambridge University Press

Williams, B., (1985a), *Ethics and the Limits of Philosophy*, Cambridge MA: Harvard University Press

Williams, B., (1985b), 'Ethics and the Fabric of the World', in: Honderich, T. (ed.), *Morality and Objectivity*, London, Routledge

Williams, B., (1995a), 'Internal Reasons and the Obscurity of Blame', in: *Making Sense of Humanity*, Cambridge, Cambridge University Press

- Williams, B., (1995b), 'Replies', in: Altham, J.E.J., R. Harrison (eds.), *World, Mind and Ethics, Essays on the Ethical Philosophy of Bernard Williams*, Cambridge: Cambridge University Press
- Williams, B., (2001), 'Postscript: Some Further Notes on Internal and External Reasons', in: Millgram, E. (ed.), *Varieties of Practical Reasoning*, Cambridge MA: MIT Press
- Williamson, T., (2007), *The Philosophy of Philosophy*, Oxford: Blackwell Publishing
- Wittgenstein, L., (1965), 'Lecture on Ethics', in: *Philosophical Review* **74**: 3-12
- Wright, C., (1996), 'Truth in Ethics', in: Hooker, B. (ed.), *Truth in Ethics*, Oxford: Blackwell
- Zangwill, N., (2012), 'Rationality and Moral Realism', in: *Ratio* **23(3)**: 345-64